147
Structure and Bonding

Series Editor: D. M. P. Mingos

Editorial Board:
F. A. Armstrong · P. Day · X. Duan · L. H. Gade
K. R. Poeppelmeier · G. Parkin · J.-P. Sauvage ·
M. Takano

For further volumes:
http://www.springer.com/series/430

Structure and Bonding

Series Editor: D. M. P. Mingos

Recently Published and Forthcoming Volumes

Electron Density and Chemical Bonding II
Volume Editor: Dietmar Stalke
Vol. 147, 2012

Electron Density and Chemical Bonding I
Volume Editor: Dietmar Stalke
Vol. 146, 2012

Structure-Property Relationships in Non-Linear Optical Crystals II
Volume Editors: Xin-Tao Wu, Ling Chen
Vol. 145, 2012

Structure-Property Relationships in Non-Linear Optical Crystals I
Volume Editors: Xin-Tao Wu, Ling Chen
Vol. 144, 2012

Molecular Electronic Structures of Transition Metal Complexes II
Volume Editors: D. M. P. Mingos, Peter Day, Jens Peder Dahl
Vol. 143, 2012

Molecular Electronic Structures of Transition Metal Complexes I
Volume Editors: D. M. P. Mingos, Peter Day, Jens Peder Dahl
Vol. 142, 2012

Fuel Cells and Hydrogen Storage
Volume Editors: Andrew Bocarsly, D. M. P. Mingos
Vol. 141, 2011

Zintl Ions
Principles and Recent Developments
Volume Editor: Thomas F. Fässler
Vol. 140, 2011

Zintl Phases
Principles and Recent Developments
Volume Editor: Thomas F. Fässler
Vol. 139, 2011

Inorganic 3D Structures
Volume Editor: Angel Vegas
Vol. 138, 2011

Molecular Catalysis of Rare-Earth Elements
Volume Editor: Peter W. Roesky
Vol. 137, 2010

Metal-Metal Bonding
Volume Editor: Gerard Parkin
Vol. 136, 2010

Functional Phthalocyanine Molecular Materials
Volume Editor: Jianzhuang Jiang
Vol. 135, 2010

Data Mining in Crystallography
Volume Editors: Hofmann, D. W. M., Kuleshova, L. N.
Vol. 134, 2010

Controlled Assembly and Modification of Inorganic Systems
Volume Editor: Wu, X.- T.
Vol. 133, 2009

Molecular Networks
Volume Editor: Hosseini, M. W.
Vol. 132, 2009

Molecular Thermodynamics of Complex Systems
Volume Editors: Lu, X., Hu, Y.
Vol. 131, 2009

Contemporary Metal Boron Chemistry I
Volume Editors: Marder, T. B., Lin, Z.
Vol. 130, 2008

Recognition of Anions
Volume Editor: Vilar, R.
Vol. 129, 2008

Liquid Crystalline Functional Assemblies and Their Supramolecular Structures
Volume Editor: Kato, T.
Vol. 128, 2008

Electron Density and Chemical Bonding II

Theoretical Charge Density Studies

Volume Editor: Dietmar Stalke

With contributions by

M. Chodkiewicz · B. Dittrich · B. Engels · R.F. Fink · S. Fux · C. Gatti · J. Henn · D. Jayatilaka · T. Koritsanszky · K. Meindl · M. Reiher · T. Schirmeister · T.C. Schmidt · A. Volkov

Springer

Editor
Dietmar Stalke
Universität Göttingen
Institut für Anorganische Chemie
Tammannstraße 4
Göttingen
Germany

ISSN 0081-5993 ISSN 1616-8550 (electronic)
ISBN 978-3-642-30807-9 ISBN 978-3-642-30808-6 (eBook)
DOI 10.1007/978-3-642-30808-6
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012939135

© Springer-Verlag Berlin Heidelberg 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply , even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Series Editor

Prof. D. Michael P. Mingos

Inorganic Chemistry Laboratory
Oxford University
South Parks Road
Oxford OX1 3QR, UK
michael.mingos@st-edmund-hall.oxford.ac.uk

Volume Editor

Dietmar Stalke
Universität Göttingen
Institut für Anorganische Chemie
Tammannstraße 4
Göttingen
Germany

Editorial Board

Prof. Fraser Andrew Armstrong

Department of Chemistry
Oxford University
Oxford OX1 3QR
UK

Prof. Peter Day

Director and Fullerian Professor
of Chemistry
The Royal Institution of Great Britain
21 Albermarle Street
London W1X 4BS, UK
pday@ri.ac.uk

Prof. Xue Duan

Director
State Key Laboratory
of Chemical Resource Engineering
Beijing University of Chemical Technology
15 Bei San Huan Dong Lu
Beijing 100029, P.R. China
duanx@mail.buct.edu.cn

Prof. Lutz H. Gade

Anorganisch-Chemisches Institut
Universität Heidelberg
Im Neuenheimer Feld 270
69120 Heidelberg, Germany
lutz.gade@uni-hd.de

Prof. Dr. Kenneth R. Poeppelmeier

Department of Chemistry
Northwestern University
2145 Sheridan Road
Evanston, IL 60208-3133
USA
krp@northwestern.edu

Prof. Gerard Parkin

Department of Chemistry (Box 3115)
Columbia University
3000 Broadway
New York, New York 10027, USA
parkin@columbia.edu

Prof. Jean-Pierre Sauvage

Faculté de Chimie
Laboratoires de Chimie
Organo-Minérale
Université Louis Pasteur
4, rue Blaise Pascal
67070 Strasbourg Cedex, France
sauvage@chimie.u-strasbg.fr

Prof. Mikio Takano

Institute for Integrated Cell-Material
Sciences (iCeMS)
Kyoto University
Yoshida Ushinomiya-cho
Sakyo-ku
Kyoto 606-8501
Japan

Structure and Bonding
Also Available Electronically

Structure and Bonding is included in Springer's eBook package *Chemistry and Materials Science*. If a library does not opt for the whole package the book series may be bought on a subscription basis. Also, all back volumes are available electronically.

For all customers who have a standing order to the print version of *Structure and Bonding*, we offer the electronic version via SpringerLink free of charge.

If you do not have access, you can still view the table of contents of each volume and the abstract of each article by going to the SpringerLink homepage, clicking on "Chemistry and Materials Science," under Subject Collection, then "Book Series," under Content Type and finally by selecting *Structure and Bonding*.

You will find information about the

– Editorial Board
– Aims and Scope
– Instructions for Authors
– Sample Contribution

at springer.com using the search function by typing in *Structure and Bonding*.

Color figures are published in full color in the electronic version on SpringerLink.

Aims and Scope

The series *Structure and Bonding* publishes critical reviews on topics of research concerned with chemical structure and bonding. The scope of the series spans the entire Periodic Table and addresses structure and bonding issues associated with all of the elements. It also focuses attention on new and developing areas of modern structural and theoretical chemistry such as nanostructures, molecular electronics, designed molecular solids, surfaces, metal clusters and supramolecular structures. Physical and spectroscopic techniques used to determine, examine and model structures fall within the purview of *Structure and Bonding* to the extent that the focus

is on the scientific results obtained and not on specialist information concerning the techniques themselves. Issues associated with the development of bonding models and generalizations that illuminate the reactivity pathways and rates of chemical processes are also relevant.

The individual volumes in the series are thematic. The goal of each volume is to give the reader, whether at a university or in industry, a comprehensive overview of an area where new insights are emerging that are of interest to a larger scientific audience. Thus each review within the volume critically surveys one aspect of that topic and places it within the context of the volume as a whole. The most significant developments of the last 5 to 10 years should be presented using selected examples to illustrate the principles discussed. A description of the physical basis of the experimental techniques that have been used to provide the primary data may also be appropriate, if it has not been covered in detail elsewhere. The coverage need not be exhaustive in data, but should rather be conceptual, concentrating on the new principles being developed that will allow the reader, who is not a specialist in the area covered, to understand the data presented. Discussion of possible future research directions in the area is welcomed.

Review articles for the individual volumes are invited by the volume editors.

In references *Structure and Bonding* is abbreviated *Struct Bond* and is cited as a journal.

Impact Factor in 2010: 4.659; Section “Chemistry, Inorganic & Nuclear”:
Rank 4 of 43; Section “Chemistry, Physical”: Rank 25 of 127

Foreword

The field of charge density determination and analysis provides a rare case of fruitful long-term symbiosis between theoretical and experimental developments. This is probably due to a temporal coincidence, since the introduction of the Hansen–Coppens pseudoatom model in 1978, which paved the way toward modern experimental achievements, was coetaneous to a renewed interest in the charge density itself as a theoretical vehicle to chemical bonding issues, as pioneered by the topological approach developed by R. Bader. Some 30 years after these initial breakthroughs, the accuracy of experimentally derived charge densities may, in favorable cases, rival that obtained with the most expensive quantum mechanical simulations. On the ab initio frontier, a wealth of descriptors have been introduced based on the electron density (or low-order reduced density matrices) that offer bonding, even energetic insights using a framework that has been enthusiastically adopted by the X-ray charge density community. Communication has been fluent on both ends. Thus, useful indices not directly accessible from the charge density, like electron localization functions, have been approximately expanded in terms of the density and its derivatives. Similarly, sloppy concepts like the pseudoatom have benefited from the space partitioning techniques developed by theoreticians.

If this partnership survives, the golden charge density era we are living in faces a brilliant future. Newly derived global (i.e., domain integrated) descriptors provide absolute (i.e., free of references) measures of concepts that range from bond orders to total energies, passing through chemical reactivity and even aromaticity measures. If they are successfully mapped onto the charge density space in the near future, we may envision a widened field that would allow access to all sort of physical, thermodynamic, and chemical properties from charge density experiments.

This volume focuses on state-of-the-art, theoretically oriented advancements in the field and offers six very different, yet important contributions. In the first chapter, Koritsanszky and coworkers revisit the pseudoatom model, which together with the maximum entropy method constitutes the core of the discipline. Attention is paid to the definition of the radial functions used in the multipole expansion of the density, which are obtained here by projecting quantum-chemically derived

Hirshfeld (or stockholder) atoms. The performance, transferability, and applicability of the procedure are also critically analyzed. The next two chapters analyze two important issues related to the quality of experimental densities. Obtaining reliable physical properties from them has led to bitter debates in the last decade, particularly as regards the magnitude of environmentally induced dipole moment enhancements. This subject is tackled by Dittrich and Jayatilaka, who compare theoretically and experimentally determined dipole moments using multipole, Hirshfeld, and X-ray constrained wave function refinements. The trustability of theoretically obtained densities and Laplacians, as well as their use as QSAR descriptors for inhibitor drugs both in the case of weak and strong (covalent) drug–enzyme interactions is studied in the chapter written by Engels and coworkers. Another issue that has deserved much attention in recent times is the role of relativistic effects on the charge density. It provides a very interesting case where the experiment, fully relativistic by definition, may help choose among a plethora of competing theoretical models, at least until fully relativistic many electron calculations become standardly available. The subtleties of relativistic electron densities, including correlation effects, are beautifully presented by Fux and Reiher. Chapter 5 is devoted to the basics of residual density analysis. As Meindl and Henn show, this tool which detects and models structures in residual maps may serve several purposes in charge density refinement and modeling. The final chapter by Gatti is a very good example of how a clever descriptor based on the density, in this case the source function, may hold clues to many electron properties like bond orders and π-conjugation, thus opening new avenues.

The contents of this volume show that there is still much room to be explored in the field. Hopefully, some of its readers will find novel solutions to the problems here examined.

Oviedo
2011

A. Martín Pendás

Preface

The electron density distribution is the ultimate observable for determining and interpreting the properties of matter. Analysing the electron density and assigning its features to properties will inevitably provide routes to rationally modifying the desired properties of matter in a predictable manner. Fortunately, the electron density may be analysed in two different and independent ways: firstly from quantum theory which leads to the molecular wave functions and secondly from the diffraction experiment. Therefore, ambiguities possibly introduced from model bias may be minimized by reference to both sets of data. Results obtained from X-ray diffraction which are subject to experimental errors may be validated by those obtained from theoretical methods. Conversely, new theoretical approaches inherently limited by computer power and memory size can be bench-marked using data from diffraction experiments.

The collection of review articles published in Volumes 146 and 147 of the Structure and Bonding Series provides a state-of-the-art overview of the experimental and theoretical determination of charge densities written by leaders in the field. We hope that their insights will motivate more scientists to take advantage of the approach.

Göttingen
April 2012

Dietmar Stalke

Contents

New Directions in Pseudoatom-Based X-Ray Charge Density Analysis 1
Tibor Koritsanszky, Anatoliy Volkov, and Michal Chodkiewicz

Reliable Measurements of Dipole Moments from Single-Crystal Diffraction Data and Assessment of an In-Crystal Enhancement 27
B. Dittrich and D. Jayatilaka

Challenging Problems in Charge Density Determination: Polar Bonds and Influence of the Environment 47
Bernd Engels, Thomas C. Schmidt, Carlo Gatti, Tanja Schirmeister, and Reinhold F. Fink

Electron Density in Quantum Theory 99
Samuel Fux and Markus Reiher

Residual Density Analysis 143
Kathrin Meindl and Julian Henn

The Source Function Descriptor as a Tool to Extract Chemical Information from Theoretical and Experimental Electron Densities 193
Carlo Gatti

Index 287

Struct Bond (2012) 147: 1–26
DOI:10.1007/430_2010_32
© Springer-Verlag Berlin Heidelberg 2010
Published online: 8 December 2010

New Directions in Pseudoatom-Based X-Ray Charge Density Analysis

Tibor Koritsanszky, Anatoliy Volkov, and Michal Chodkiewicz

Abstract This chapter revisits critical aspects of pseudoatom interpretation of X-ray Bragg diffraction data to derive the crystalline electron density experimentally. The main focus is on the radial basis functions in the nucleus-centered multipole expansion of the model density. Based on the direct-space projection of the stockholder-partitioned quantum-chemical molecular density onto nucleus-centered spherical harmonics, "bonded-atom" radial functions are derived, in terms of which the original density can be reconstructed. The method allows for a quantitative access of errors introduced by restrictions on the pseudoatom radial functions and for designing more adequate basis sets than those routinely used in the standard model. The applicability of the upgraded multipole formalism to X-ray charge density analysis is demonstrated through pseudoatom fitting of simulated and experimental data. The chemical transferability of stockholder pseudoatoms is analyzed through small-molecule model studies. A protocol to build a stockholder-atom library and the feasibility of simple data-refinement strategies using the databank radial functions is also discussed.

Keywords Electron density · Pseudoatom model · Radial functions · Stockholder partitioning · X-ray diffraction

Contents

1 Introduction 2
2 Static Electron Density from X-Ray Structure Factors 3
3 Multipole Expansion of the Electron Density 4
4 The Standard Pseudoatom Formalism 5
5 Pseudoatom RDFs from Molecular Densities 7
6 Absolute Error of the Pseudoatom Density 10
7 X-Ray Data Analysis with Theoretical Stockholder Pseudoatoms 11

T. Koritsanszky (✉), A. Volkov, and M. Chodkiewicz
Department of Chemistry, Computational Science Program, Middle Tennessee State University, Murfreesboro, TN 37132, USA
e-mail: tkoritsa@mtsu.edu

8 Chemical Transferability of Stockholder RDFs 15
9 Stockholder Pseudoatom Databank 18
10 Summary 20
References 22

1 Introduction

Diffraction techniques are the primary experimental sources of information on the nuclear and electronic structure of solids. There is, however, no direct route from data to the information they contain. This inversion problem necessarily involves modeling the structure, the scattering process, and the intensity detection. The reliability and completeness of information retrieved from the observations is thus subject to the adequacy of the models applied. Coherent elastic diffraction is the most developed technique. X-ray Bragg intensities are almost routinely analyzed nowadays, not only to derive crystal structures but also to extract the static electron density (ED: $\rho(\mathbf{r})$) of solids [1, 2]. Since Bragg scattering is associated with the average structure, the interpretation of these data must also include a model for decoupling electronic and nuclear motions [3, 4]. This can be conveniently done within the one-center approximation, such as the pseudoatom (PA) formalism [5, 6], which is a nucleus-centered multipole expansion of the crystalline ED. To be applicable to finite data fitting, the expansion must be finite and efficiently parameterized. These requirements pose severe restrictions on the PA radial functions (RDF). In the most popular PA model [6], the RDFs assigned to the nonspherical part of the density are in fact single Slater functions shared by all real spherical harmonics (RSH) of the same order (l). The overall good data-fitting performance of pseudoatoms and the meaningful chemical content of the corresponding fitted static ED are thus somewhat unexpected. The majority of studies report a good quantitative agreement between experimental and theoretical densities, especially for covalent bonds in light-atom molecular crystals. The agreement in the internuclear regions occurs, however, only at the expense of disagreement in the nuclear regions – a clear manifestation of the signal redistribution characteristic of the Fourier transform. The chemically oriented scientific community applies high-resolution crystallography with increasing confidence to study chemical bonding. The number and complexity of the properties that are claimed to be reliably accessible by routine application of the method have been increased significantly over the past decade [7]. The objective of PA-based X-ray charge density investigations has been shifted from semi-quantitative description of bonding features to full topological analysis [8] of the model density and related properties, whose spectrum has been extended even to energy densities [9–15]. Furthermore, recent applications target systems of increasing size and complexity that include biomacromolecules [16, 17], extended solids [18–20], framework materials containing heavy elements [21, 22], and organometalics [23–25]. The questions as to what extent this increased confidence in the method is justifiable and whether different errors indeed cancel each other favorably remain important to be addressed.

2 Static Electron Density from X-Ray Structure Factors

Within the kinematic theory, single-crystal X-ray diffraction is treated as a coherent elastic scattering of X-ray photons on electrons with periodic distribution ($\rho(\mathbf{r})$). The key property relevant to this process is the Bragg structure factor ($F(\mathbf{H})$) which is the Fourier transform of the mean electron density ($\langle\rho\rangle$) in the unit cell of volume V_c.

$$F(\mathbf{H}) = \int_{V_c} \langle\rho(\mathbf{r})\rangle e^{i2\pi\mathbf{Hr}} d\mathbf{r} = \sum_A \hat{f}_A(\mathbf{H}) t_A(\mathbf{H}) e^{i2\pi\mathbf{HR}_A} \tag{1}$$

H is the scattering vector, with integral reciprocal-axis components (Miller indices h, k and l), satisfying the Laue conditions ($\boldsymbol{Ha}^* = h$, $\boldsymbol{Hb}^* = k$, and $\boldsymbol{Hc}^* = l$) dictated by the translation symmetry of the crystal density. Assuming that the system remains in the ground electronic state during the scattering process, the averaging is taken only over the vibrational states. The right-hand side of (1) is obtained within the harmonic convolution approximation to thermal smearing [4], which requires the static crystalline density to be partitioned into "atomic" densities ($\hat{\rho}_A$), each of which rigidly follows the motion of a particular nucleus at $\mathbf{R}_A$:

$$\rho(\mathbf{r}) = \sum_A \hat{\rho}_A(\mathbf{r}_A); \quad \mathbf{r}_A = \mathbf{r} - \mathbf{R}_A \tag{2}$$

The atomic scattering factor ($\hat{f}_A$) and the temperature factor (t_A) are, respectively, the Fourier transform of $\hat{\rho}_A$ and the Gaussian probability density function ($P(\mathbf{u}_A)$) of harmonic nuclear displacements with respect to the equilibrium configuration ($\mathbf{u}_A = \mathbf{Q}_A - \mathbf{R}_A$):

$$P(\mathbf{u}_A) = (2\pi)^{-3/2} |\mathbf{U}_A|^{-1/2} e^{-\frac{1}{2}\mathbf{u}_A' \mathbf{U}_A^{-1} \mathbf{u}_A} \tag{3}$$

The components of the mean-square displacement amplitude matrix ($\mathbf{U}_A = \langle \mathbf{u}_A \mathbf{u}_A' \rangle$) are referred to as the atomic (anisotropic) displacement parameters (ADP).

Atomicity is a driving concept of X-ray diffraction analysis. It is invoked at different stages of the structure elucidation, starting from initial phase assignment, through structure refinement, to accurate modeling of solid-state charge distributions. The conventional structure refinement is based on the promolecule density ($\rho^o(\mathbf{r})$), which is the superposition of spherical atomic densities (ρ_A^o) derived from quantum-chemical calculations on isolated atoms:

$$\rho^o(\mathbf{r}) = \sum_A \rho_A^o(\mathbf{r}_A) \tag{4}$$

Different one-center density models used for X-ray data fitting can be considered as extensions of the promolecule to account for local deformations due to chemical bonding.

3 Multipole Expansion of the Electron Density

Since the atomic Hamiltonian commutes with the square of the total angular momentum operator, the physical density of an atom in a well-defined angular momentum state (L, M) has a well-defined RSH ($y_{lm}(\vartheta, \phi) = y_{lm}(\Omega)$) content [26]:

$$\rho(\mathbf{r}) = \sum_{l}^{L} \rho_{2l,0}(r) y_{2l,0}(\Omega) \tag{5}$$

Due to the RSHs forming a complete orthogonal basis, the product of any two can be written as a linear combination of a finite number of RSHs. Thus, the density of an atom obtained within the orbital approximation also has a finite multipole expansion, as best seen for a single-determinant wave function composed of one-electron spin-functions. Within the LCAO-MO approximation, the molecular density can be decomposed into one- and two-center orbital product [27]. The multipole content of each one-center density, just like for an atom, is uniquely determined by the orbital basis.

The many-centered finite multipole expansion of the molecular or crystalline ED was introduced by Stewart [5]. In his formalism, the total density ($\rho(\mathbf{r})$) is decomposed into *pseudoatoms*:

$$\rho(\mathbf{r}) = \sum_{A} \sum_{l,m}^{L} \rho_{lm}^{\mathrm{A}}(r_A) y_{lm}(\Omega_A) \tag{6}$$

whose RDFs (ρ_{lm}^{A}) can in principle be derived by minimization of the mean-square residual (MSR: χ^2) between the target and model densities:

$$\chi^2 = \int \left(\rho(\mathbf{r}) - \sum_{A} \sum_{l,m}^{L} \rho_{lm}^{A}(r_A) y_{lm}(\Omega_A) \right)^2 \mathrm{d}\mathbf{r} = \int_0^\infty \chi_0^2(r) r^2 \mathrm{d}r \tag{7}$$

χ_0^2 is a positive definite radial MSR obtained by integrating χ^2 over the angular variables. Its minimization with respect to each RDF leads to a set of inhomogeneous linear equations:

$$\rho_{lm}(r_A) = \left\langle \rho \mid y_{lm}^{A} \right\rangle_{\Omega} = \rho_{lm}^{A}(r_A) + \sum_{B \neq A} \sum_{\lambda\mu}^{L} \rho_{\lambda\mu}^{B}(r_B) \left\langle y_{\lambda\mu}^{B} \mid y_{lm}^{A} \right\rangle_{\Omega} \tag{8}$$

The molecular RDF ($\rho_{lm}(r_A)$) of the A-centered multipole expansion (l.h.s.) at any radial grid point is given as a linear combination of pseudoatom RDFs (r.h.s.) taken at the same radial grid point. The projection, driven by the angular overlap integral of RSHs $\left(\langle y^B_{\lambda\mu} \,|\, y^A_{lm}\rangle_\Omega\right)$, yields "exact" molecular moments up to L, without necessarily reproducing the total density exactly. The procedure leads to delocalized RDFs, each depending on the nuclear geometry and the level of expansion (L) at each center [28].

4 The Standard Pseudoatom Formalism

Since the above derivation of RDFs from a finite set of X-ray structure factors is not feasible, a parameterized version of the expansion has been developed. In the popular Hansen–Coppens formalism [6] (referred to here as the *standard* or HC-pseudoatom model (HC-PA), the RDFs are predefined as follows:

$$\rho_{00} = \rho_{\mathrm{C}} + P_{00}\rho_{\mathrm{V}}(\kappa r), \quad \rho_{lm} = P_{lm}S_l(\kappa' r), l > 0; \quad -l \le m \le +l \tag{9}$$

The monopole part is taken as a combination of the spherical Hartree–Fock frozen-core (ρ_{C}) and normalized valence densities [29] (ρ_{V}) of the isolated atom, while those in the deformation term (S_l) are density normalized Slater functions:

$$S_l = \frac{{\eta_l}^{n_l+3}}{(n_l+2)!} r^{n_l} e^{-\eta_l r} \tag{10}$$

The parameters (n_l and η_l) are deduced from energy-optimized single-zeta Hartree–Fock (HF) atomic wave functions [30]. The radial deformation of the valence density is accounted for by the expansion–contraction variables (κ and κ'). The RSHs are density normalized and expressed in nucleus-centered local frames:

$$\int_\Omega \left|d_{0,m}(\Omega)\right| d\Omega = 1 \quad \text{and} \quad \int_\Omega \left|d_{l>0,m}(\Omega)\right| d\Omega = 2 \tag{11}$$

The Fourier transform of the pseudoatom ED enters into the structure-factor expression (1) as a complex but analytic scattering factor:

$$f(\mathbf{H}) = f_{\mathrm{C}}(H) + P_{\mathrm{V}} f_{\mathrm{V}}(H/\kappa) + 4\pi \sum_l i^l J_l(H/\kappa') \sum_{m=-l}^{+l} P_{lm} d_{lm}(\Omega_H) \tag{12}$$

where, f_{C} and f_{V} are the Fourier transform of ρ_{C} and ρ_{V}, respectively, J_l is the l-order Fourier–Bessel transform of S_l, and Ω_H encompasses the local polar angles of the scattering vector ($\boldsymbol{H}$). The static ED parameters, P_{lm}, κ and κ', are optimized, together with the nuclear positions and ADPs, in a least-squares (LS) refinement against a set of measured structure-factor amplitudes.

The HC formalism was developed with the sole purpose of extracting the static ED of crystalline materials from high-resolution X-ray data. The experimental effort involved was technically demanding and much more time consuming than nowadays. Since the accessible data are limited in resolution, the analysis requires an economic model which performs reasonably well even with a small number of fitting parameters (at least an order of magnitude smaller than the number of data points to be fitted). The HC-PA effectively compromises between flexibility and locality – an obvious reason for its success. The formalism has survived a period of over 30 years during which it has been challenged by data of increased quality and quantity. The first point to be emphasized is that the HC-PA is structure independent, that is, no parameters related to the local connectivity or bonding of the atom are included in its analytic expression (9). One of its obvious limitations is the frozen-core approximation, that is, the core polarization, which plays an important role in balancing electrostatic forces [31], is not accounted for directly. This effect is expected to be pronounced for heavier elements, since the scattering power of their atoms is dominated by the core electrons. The radial signals due to core polarization are extremely sharp, localized in the vicinity of the nuclei, and thus manifest themselves mainly in the high-order reflections, for which, however, the experimental signal-to-noise ratio is low. The extent to which the frozen-core approximation biases the valence density parameters is thus not readily, if at all, accessible by X-ray diffraction.

The most important sources of bias are the restrictions imposed on the valence RDFs. The expansion/contraction of ρ_V upon charge gain/loss is accounted for by a single scaling parameter (κ). Simple chemical arguments, however, suggest that the sub-shells are not uniformly deformed upon bond formations. It is also evident that single Slater functions are not flexible enough to describe radial deformations in molecules. Moreover, their direct construction from the ground-state minimal-basis atomic orbitals is ambiguous. For example, no odd-order and hexadecapolar RDFs can be assigned to a quadrupolar atom (*s*,*p*-valence shell) without mixing *s* and *p* orbitals according to some hybridization scheme. For atoms with outermost *d* subshell, the *s* and *p* contributions from the same shell can either be included (giving rise to all RDFs up to $l = 4$, but with mixed orbital contributions) or kept frozen (leading to S_2 and S_4, but with pure *d*-orbital contribution). The RDF set deduced in such a way is unbalanced, since ρ_V is derived from the atomic wave functions of near-HF-limit, while the deformation RDFs are constructed from single Slater orbitals. Furthermore, these functions are *m*-independent, unlike those in the general expansion (6). This means that poles with the same angular (l) but different magnetic quantum number (m) share the same S_l. This restriction prohibits distinguishing between *p*-orbitals involved in π and σ bonding, which was recognized as an important requirement for an adequate basis set (split-valence) already at the early stage of developments of routine molecular orbital (MO) calculations. Last but not least, the same set of RDFs is assigned to centers with the same atomic number, irrespective of their local bonding situation.

The adequacy of the HC formalism and the physical content/reliability of the ED projected onto the model from a finite set of structure factors are frequently raised

issues, which have directly motivated several investigations [32–34] and which are addressed to a great extent in almost all X-ray charge density studies. The validation includes the comparison of a selected experimental result (a particular LS solution) with a theoretical one, which is usually done in terms of a visual inspection of static deformation ED maps; one for the crystalline molecule extracted from the experimental data, and one for the corresponding isolated molecule calculated from an approximate wave function at the experimental geometry. The exhausting analysis of "reference free" topological properties [35] derived by both methods has become a common practice and is included in the larger body of published works [7, 36]. This comparison is, however, more of a mandatory than a conclusive exercise, since both methods have their own limitations which forbid unambiguous tracking of the source of disagreement or, in fact, the reason for agreement. In other words, it is difficult, if not impossible, to conclude whether the discrepancies are due to artifacts or real physical effects.

Studies using synthetic data calculated via Fourier transform of theoretical densities, obtained either from periodic or isolated molecule wave functions, are becoming increasingly popular [37–39]. Such simulations can indeed reveal model inadequacies, since the data are free of errors, correctly phased, and can be generated to high resolution with or without thermal smearing. The method has been used to explore the bias in the fitted PA density due to the restricted RDFs of the standard model [40, 41] and to derive "theory-supported" PA parameters adoptable in experimental studies [42, 43]. Depending on the quality of the target density, simulations sometimes can lead to surprisingly good results for organic molecules even in terms of local (ρ_{BCP}, $\nabla^2\rho_{\mathrm{BCP}} = \lambda_1 + \lambda_2 + \lambda_3$, where λs are the principal curvatures of ρ at the Bond Critical Point: BCP) and integrated (Atoms in Molecules) topological figures. The fitted density is, however, severely biased or even meaningless if the target density is calculated using an extended basis set. An extreme example is presented in Fig. 1 which compares the direct-space difference density ($\rho_{\Psi} - \rho_{\mathrm{HC-PA}}$) with the HC-PA deformation density in the O–S–S plane of the *heptasulfur imide* molecule (S_7NH). The fitted PA density ($\rho_{\mathrm{HC-PA}}$) was obtained by a phase-restricted multipole refinement of the static ED parameters, using the default RDFs of the XD program package [44], against structure factors generated at the MP2/cc-pvtz level of theory [45–47] for the isolated molecule at the experimental geometry [48]. In spite of an excellent structure-factor fit ($R = 0.24\%$), the absolute error map (Fig. 1a) exhibits features that are comparable to those found on the deformation ED map (Fig. 1b), making thus such an interpretation of bonding effects meaningless.

5 Pseudoatom RDFs from Molecular Densities

In view of the foregoing comments, efforts toward upgrading the standard HC-PA model must focus on the development of RDFs from molecular rather than atomic densities. A feasible approach, which adopts projection (7) but avoids its

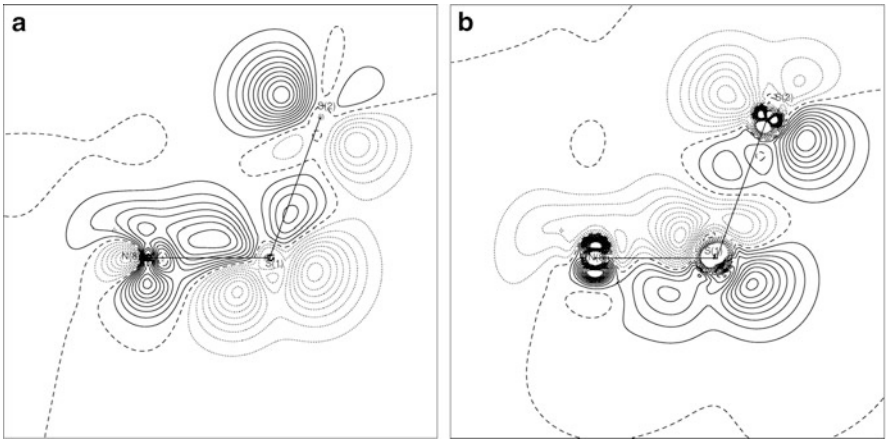


Fig. 1 (**a**) Static difference electron density map ($\rho_\Psi - \rho_{\text{HC-PA}}$) calculated for the *heptasulfur imide* molecule in the N–S–S plane. The pseudoatom density was obtained via fitting the parameters of the HC-PA model to MP2/cc-pvtz structure factors. (**b**) Deformation electron density map ($\rho_\Psi - \rho^0$) in the same plane. Contour levels are 0.05 eÅ^{-3}. *Solid*, *dashed*, and *dotted lines* are positive, zero, and negative, respectively

difficulties, is provided by a two-step procedure [49]; the target molecular ED (ρ_Ψ) is first fragmented into fuzzy atomic densities (ρ_A),

$$\rho_\Psi = \sum_k n_k \Psi_k^2 \Rightarrow \rho_\Psi = \sum_A \rho_A \tag{13}$$

followed by the projection of each of these units onto RSHs:

$$R_{lm}^A(r_A) = \frac{\int \rho_A(\mathbf{r}) d_{lm}^A d\Omega_A}{\int \rho_A(\mathbf{r}) d_{lm}^A d\mathbf{r}_A} \tag{14}$$

The summation in (13) is over the number of occupied MOs (Ψ_k), and the integral in the denominator of (14) gives rise to the multipole populations (Q_{lm}^A). The normalized RDFs obtained in such a way correspond to "bonded-atoms" (derived from molecular rather than atomic wave function) and are uncorrelated because of the orthogonality of the RSHs at the same center. The integrals are calculated numerically, on a fine grid of the radial coordinate measured from the nucleus, to allow for the use of any type of molecular density and partitioning. Each numerical RDF is fitted with a combination of Slater primitives (10).

Different fuzzy-type decomposition schemes are expected to yield similar RDFs because the MO density is dominated by the one-center orbital products. The results discussed here are based on the *stockholder* partitioning method [50], which

optimally compromises between locality and transferability [51]. This scheme fragments the total ED (ρ_Ψ) into densities of "chemical atoms" (ρ_A) in proportion to their investments into the promolecule:

$$\rho_A(\mathbf{r}_A) = \frac{\rho_A^{\mathrm{o}}(\mathbf{r}_A)}{\rho^{\mathrm{o}}(\mathbf{r})}\rho_\Psi(\mathbf{r}); \quad \sum_A \rho_A(\mathbf{r}_A) = \rho_\Psi(\mathbf{r}) \tag{15}$$

The combination of (14) and (15) yields the *unrestricted* stockholder pseudoatoms (U-SPA), since their superposition can in principle reproduce the target molecular density within an arbitrary accuracy that depends only on the level of expansion (L):

$$\rho_\Psi \cong \sum_A \sum_{lm}^{L} Q_{lm}^A R_{lm}^A d_{lm}^A \tag{16}$$

where, each RHS is augmented by an individual (m-dependent) RDF, just like in the general pseudoatom expression (6). In order to mimic the standard model, m-independent RDFs (R_l) can be derived using the technique of principal component analysis. The transformation between the two sets of RDFs can be achieved by an LS-projection:

$$\sum_{m=-l}^{+l} Q_{lm} R_{lm} y_{lm} \Leftrightarrow R_l \sum_{m=-l}^{+l} q_{lm} d_{lm} \tag{17}$$

Taking the m-independent functions as linear combinations of m-dependent functions,

$$R_l = \sum_{m=-l}^{+l} C_{lm} R_{lm} \tag{18}$$

the mixing coefficients (C_{lm}) are obtained as the components of the eigenvector corresponding to the largest eigenvalue of the overlap matrix formed by the R_{lm} set. This approach leads to the *best* possible m-independent functions in the LS sense. The corresponding model is referred here as the *restricted* stockholder pseudoatom (R-SPA). The analytic representation of the SPA-RDFs requires several Slater primitives, the actual number of which depends on the accuracy desired and on the azimuthal quantum number (l). We have developed a fitting procedure that literally "learns the shape" of these functions and derives the optimal number of Slater primitives required for each R_{lm} or R_l. The Slater exponents are preoptimized using an evolutionary algorithm. In order to arrive at a model closest to the HC-PA, each R_l is to be fitted with only a single Slater function ($\bar{S}_l$), yielding the *optimized* or 1STO-SPA.

6 Absolute Error of the Pseudoatom Density

The methods outlined above allows for the evaluation of the absolute error in the pseudoatom ED corresponding to all three models (U-, R-, and 1STO-SPA). In what it follows, we show how the total density of *α-glycine* can be reconstructed with different RDFs. The MO density was calculated at the experimental geometry [52] with the ADF program package [53] using BLYP level of theory [54, 55] and a quadruple-zeta polarized basis set (QZ4P) [56]. The projection of m-dependent RDFs (R_{lm}) onto m-independent ones (R_l) was also done numerically to avoid accuracy loss due to fitting.

Figure 2 displays residual EDs calculated in the plane of the carboxylate group with reference to the target density at different level of expansions ($L = 0$–4) using the three sets of RDFs; R_{lm} (Fig. 2a), R_l (Fig. 2b), and $\bar{S}_l$ (Fig. 2c).

It is clear that the U-SPA model terminated at the hexadecapolar level ($L = 4$) can reconstruct the target density within an overall accuracy of 0.05 e/Å^3. The maximum/minimum residual density of 0.043/−0.047 e/Å^3 is located in the close vicinity of the oxygen nuclei and rapidly diminishes as the level of expansion increases. All bonding residual contours are less than 0.02 e/Å^3, which can be further reduced if the multipole expansion is extended to $L = 6$. As expected, the ED reconstructed by the R-SPAs has a lower accuracy; the residual density amounts to about 0.05 e/Å^3 in the C–O bonds, but it is much more pronounced near to the oxygen nuclei. The most important finding is that these systematic features cannot be accounted for even with the inclusion of higher-order multipoles, since the convergence stops already at the octupolar level ($L = 3$). The 1STO-SPA model appears to be a relatively modest representation of the true ED, especially in bonds of π character. We find residual ED peaks of 0.15 e/Å^3 in the C–O bonds and maximum/minimum residual EDs of 0.163/−0.235 e/Å^3 near to the oxygen nuclei.

Table 1 lists percentage errors in the Laplacian at the BCP of different bonds obtained from the reconstructed U- and R-SPA densities, as well as from the HC-PA model fitted to the structure factors corresponding to ρ_Ψ. The values are calculated with reference to $\nabla^2\rho_\psi$. While the U-SPA model (using R_{lm} up to $L = 4$) well accounts for the target density, the Laplacian is rather poorly reproduced, especially in polar bonds. A detailed analysis reveals that this is primarily due to the limit of the level of expansion; one must extend the *U-SPA* model up to $L = 18$ to obtain the "exact" Laplacian ($\nabla^2\rho_\psi$) at the C–O bond with four significant figures. Interestingly, the R-SPAs with m-independent RDFs ($L = 4$) can occasionally perform better in predicting the bond-parallel curvature (λ_3) at the BCP than the U-SPAs. The error in the principal curvatures due to the use of restricted RDFs and the error arising from the limited level of expansion appear to cancel each other. The Laplacian values obtained from the data-fitted HC density are, however, markedly less accurate, exhibit no correlation with those from the R-SPAs, and strongly depend on the details of the refinement.

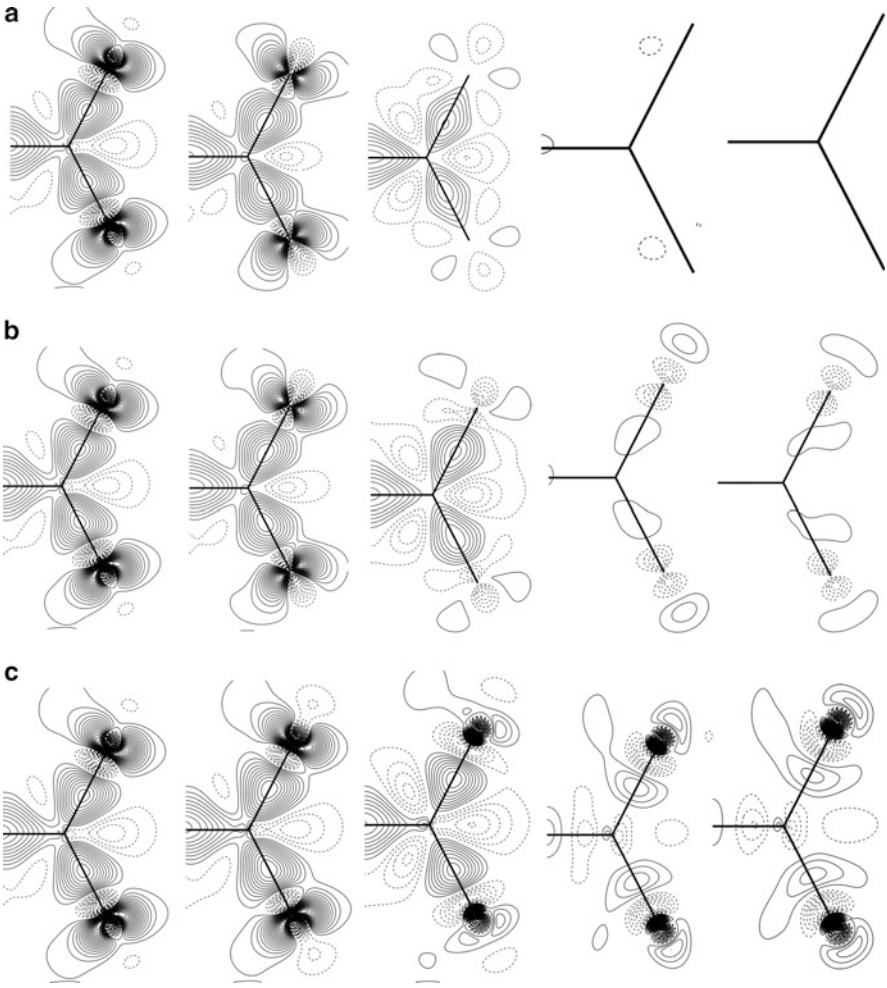


Fig. 2 Residual ED maps in the plane of the carboxylate group in α-*glycine* ($\rho_{\Psi} - \rho_{\mathrm{SPA}}$) at the B3LYP/QZ4P level). (**a**) Stockholder pseudoatom with numerical *m*-dependent RDFs; (**b**) restricted stockholder pseudoatom with *m*-independent numerical RDFs; (**c**) best-standard model with *m*-independent RDFs each fitted by a single Slater function. Contours are drawn in steps of 0.05 e/Å^3; *solid* positive, *dotted* negative. From *left* to *right*, the level of expansion is increasing from $L = 0$ to $L = 4$

7 X-Ray Data Analysis with Theoretical Stockholder Pseudoatoms

The direct-space projection technique described above not only provides an efficient validation tool to access the bias in experimental properties but also helps design new data-refinement strategies. A possible protocol for extracting the "left-over"

Table 1 Percentage error of the Laplacian at the BCP for different bonds of α-*glycine* calculated from the three pseudoatom densities at the hexadecapolar level of expansion. The values are taken relative to those obtained for ρ_Ψ (BLYP/QZ4P). Units are in *e* and Å. The HC-PA density was obtained by fitting the structure-factors corresponding to ρ_Ψ

Bond	U-SPA	R-SPA	HC-PA
C(4)-O(1)	26.8	20.2	31.8
C(4)-O(2)	24.3	15.6	27.4
C(4)-C(5)	4.19	3.50	28.0
C(5)-N(3)	3.17	3.17	2.40
N(3)-H	12.1	19.2	52.2
C(5)-H	11.2	5.20	34.9

information in experimental data of crystals of small-to-medium size molecules starts where the HC-model-based refinement ends. The first step is to project a quantum-chemical molecular density (ρ_Ψ) calculated at the HC-PA experimental geometry onto SPAs. The refinement is then repeated using analytic bonded-atom scattering factors obtained in such a way. One can initially fix the static density while adjusting only the ADPs. Since the theoretical SPA exhibits the correct behavior even near to its nucleus, the bias in the ADPs is most likely to be reduced. There is also a good chance to refine the ADPs and/or the position of H-atoms. Furthermore, this restricted protocol has the potential to identify anharmonic sites and allows for using thermal motion models beyond the harmonic approximation – an important feature that avoids the known difficulties associated with the joint adjustment of higher-order displacement amplitudes and multipole populations [57]. The final step is to refine the SPA density parameters with the hope of revealing crystal-field effects by inspecting the fitted density relative to ρ_Ψ.

Having reanalyzed several small-molecule data sets, we tend to conclude that the U-SPA model-based refinement is extremely stable. In all cases studied, a smooth and fast convergence was achieved, even for the radial screening parameters (κ') which usually exhibit a poor convergence when the standard HC-PA model is used. Furthermore, it was almost always possible to refine individual parameters (κ'_l) for individual SPA-RDFs of the non-hydrogen atoms. A related finding is that the final parameter estimates are fairly independent on the details of refinement. This robustness of fitting makes it unnecessary to monitor the "response of the data" to the selection of and/or the constraints between the variables to be refined. The parameter estimates are simply much less correlated than those obtained via the HC model using the same data set. The data "filtering capability" of the U-SPA model is also markedly better; the parameters are less biased because the model absorbs less systematic errors.

The comparative X-ray charge density study of *Bullvalene* ($C_{10}H_{10}$) presents an example which can be termed as typical, based on the limited number of preliminary investigations we pursued using medium-quality data sets. The HC-PA analysis of the 100 K serial data [58] leads to BCP properties that well correlate with the theoretical results calculated for the isolated molecule for the experimental geometry at the MP2/cc-pVTZ level. The molecule is a conjugated cage system exhibiting

three types of C–C bonds (single, double, and bent), but with intermediate bond-orders due to extended π-electron delocalization. The refinement of ADPs, using the U-SPAs projected from the MP2 density, yields basically the same reliability factor as that obtained using the HC-PAs ($R_w = 2.32\%$) with 63 extra-adjusted parameters, but a slightly better residual and a considerably different deformation ED map (Fig. 3), especially in the plane of the cyclopropane ring. It is to be emphasized that no density parameters were adjusted and the H-atom's positions and ADPs were fixed to their values derived by neutron diffraction. The subsequent refinement, including the ADPs of H-atoms (still with the fixed SPA density), led to thermal parameters in fair agreement with those derived from neutron data

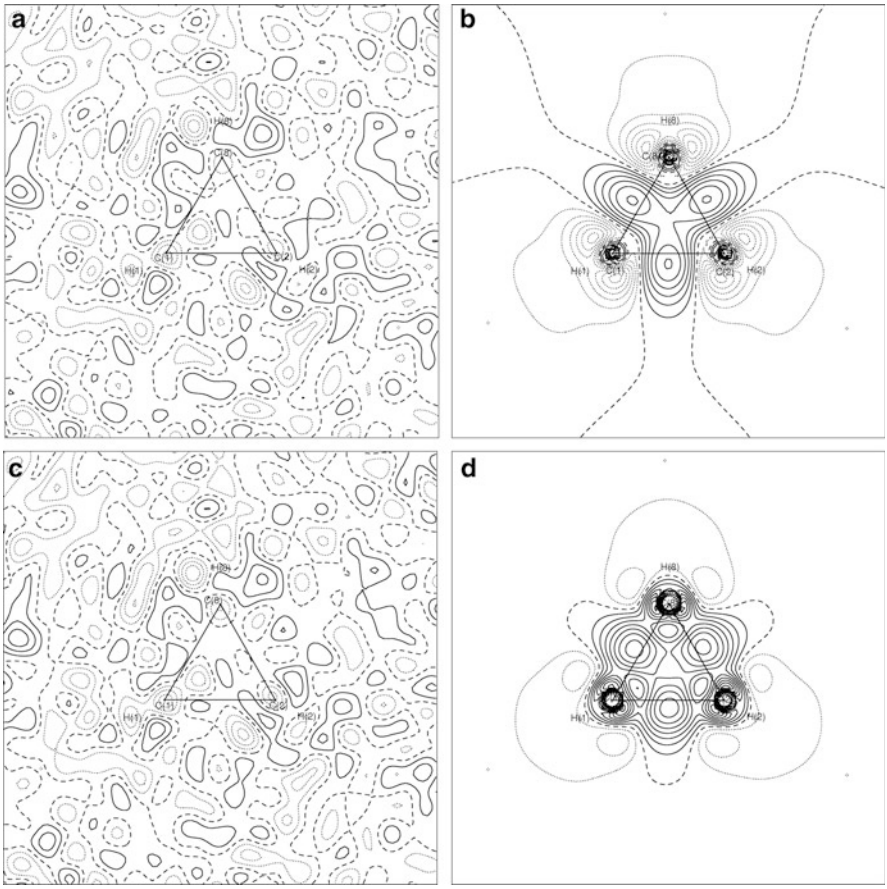


Fig. 3 (**a**, **c**) Experimental Fourier residual and (**b**, **d**) static deformation density maps in the plane of the cyclopropane ring of *Bullvalene*. *Upper row*: HC-PA refinement; *lower row*: ADP-only refinement at the same geometry using MP2/cc-pVTZ stockholder pseudoatoms. *Contour lines* are drawn at 0.05 and 0.1 eÅ^{-3} for the residual and the deformation maps, respectively. *Solid*, *dashed*, and *dotted lines* are positive, zero, and negative, respectively

collected at the same temperature. Similar results were obtained for other structures, indicating that the adequate modeling of the density ($L = 4$) can indeed help assign reasonable ADPs to the H-atom sites [59]. We also note that the all-parameter refinement of the U-SPA model brings no significant improvement in the quality of the fit or the static ED.

The analysis of experimental data of *heptasulfur imide* (Fig. 4d) leads to even more drastic differences between the theoretical (MP2/cc-pvtz) U-SPA and the fitted HC-PA model densities. Figure 4a, b depicting the residual Fourier maps obtained by the two methods and the direct-space static difference map (Fig. 4c) in the N–S–S plane of the molecule clearly demonstrates that the two markedly different models exhibit equal statistical consistency with the same data.

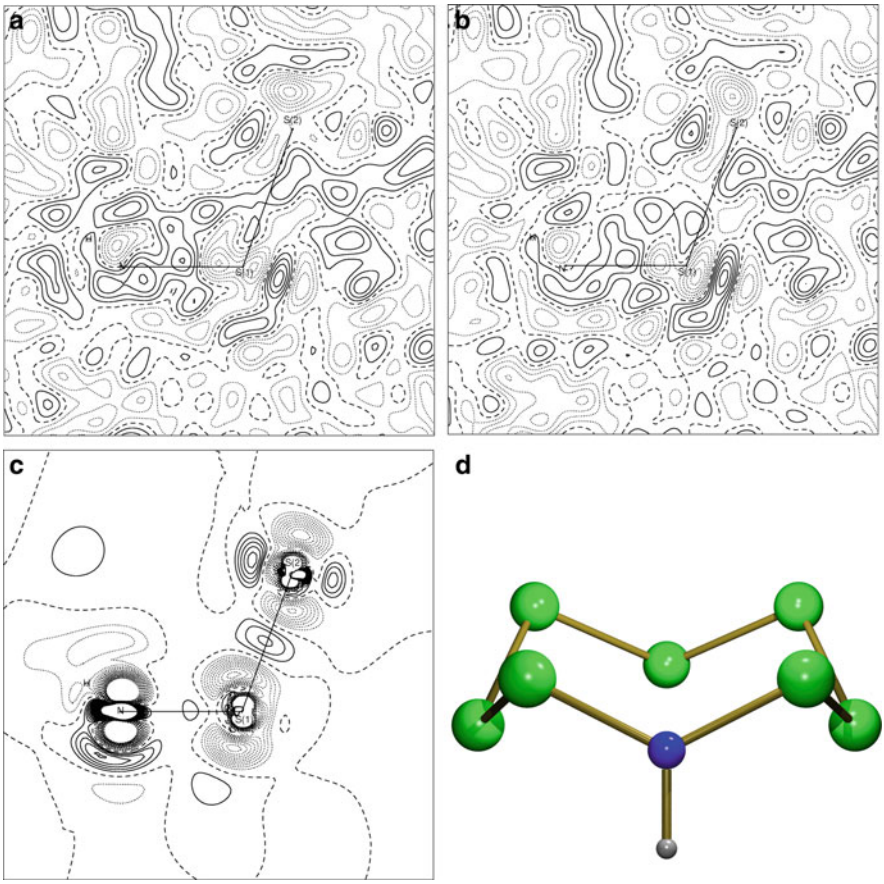


Fig. 4 Experimental Fourier (**a**, **b**) and direct-space difference (ρ_{Ψ}-$\rho_{\text{HC-PA}}$) maps (**c**) in the N–S–S plane for the *heptasulfur imide* molecule (**d**). (**a**) HC-PA refinement; (**b**) ADP-only refinement based on the MP2/cc-pVTZ U-SPAs at the same geometry. *Contour lines* are drawn at 0.05 eÅ^{-3}. *Solid*, *dashed*, and *dotted lines* are positive, zero, and negative, respectively

This observation somehow questions the reliability of residual maps for model validation. Indeed, current efforts to develop novel methods of residual analysis [60] are initiated by the recognition that neither a good fit nor the lack of visually detectable systematic features of the Fourier residual map can justify confidence in the results.

8 Chemical Transferability of Stockholder RDFs

In order to develop SPA-RDFs applicable to routine X-ray charge density analysis, it is important to explore their chemical transferability and dependence on the quantum-chemical methods and basis sets (BS) used for the calculation of ρ_Ψ. Our preliminary studies addressing the latter two issues indicate that methods including electron correlation at the MP2 or B3LYP level and using triple-zeta BSs are necessary to obtain SPA-RDFs consistent with higher-level calculations.

To study the transferability of the SPA-RDFs, a set of small organic molecules [61], composed of atoms in similar bonding environments or functional groups, were selected from the Cambridge Structural Database [62]. These include accurately determined crystal structures ($R < 5\%$) based on low-temperature X-ray or, preferably, neutron data. For X-ray-based structures, the hydrogen positions were modified by extending $X{-}H$ distances ($X = C, N, O$) to their standard neutron diffraction values [63]. All wave-function calculations were performed at MP2/aug-cc-pVTZ level of theory. The SPA-RDFs were generated and normalized via numerical three-dimensional integration using a combination of angular and radial quadratures; *Lebedev-Laikov* [64] grids for the former, while *Gauss-Chebyshev* quadrature [65] with the *Treutler-Ahlrichs M4* mapping [66] for the latter. In a typical run, three-dimensional integration was done using 200 radial and 5,810 angular points, and the functions were evaluated in the interval of ($0 \leq r \leq 2$ Å) with a grid spacing of 0.01 Å. The multipole populations and the corresponding SPA-RDFs are averaged over chemically equivalent atoms found in the target set of molecules. Two atoms are considered to be chemically equivalent if they have the same connectivity (bonded to the same type of atoms) and similar bonding situation. For the clarity of the following rather technical discussion, we introduce a compact notation; the SPA-RDF (R_{lm}) of an X-type pseudoatom ($X = H, C, N,$ and O) in a local chemical environment (E) is designated as $X_R_{lm}[E]$. The environment designation usually refers to a functional group. Two atoms of the same type but with different bonding in a group are marked with different character style in the group's formula; $O_R_{lm}[C\boldsymbol{O}OH]$, for example, refers to the SPA-RDF of the keto O-atom in the carboxyl group. The averaging led to 8, 5, 4, and 7 groups of H, C, N, and O atoms, respectively. The population analysis reveals that the stockholder scheme gives atomic charges consistent with the chemical expectation. This is best seen for the H-atoms; the monopole/bond-directed dipole population gradually decreases/increases with the increase in the electronegativity of the X atom in the X–H bond (Table 2). The deformation density of the C- and O-atoms is mainly of

Table 2 Significant multipole populations of hydrogen atoms averaged over equivalents. N_{obs} is the number of entries for each set of different H-atom types

	00	11−	10	22−	22+	30	N_{obs}
CH	0.953(7)		0.072(2)				43
CH_2	0.951(14)		0.071(4)				30
CH_3	0.957(8)		0.069(3)				21
NH	0.881(9)		0.085(3)		0.011(3)		5
NH_2 pyramidal	0.893(6)		0.084(2)	−0.013(3)			14
NH_2 planar	0.868(7)		0.091(2)		0.008(1)		20
NH_3	0.820(13)		0.089(5)				18
COOH	0.802(4)	0.012(1)	0.099(2)			−0.011(1)	5
OH	0.833(9)	0.008(2)	0.096(1)		0.009(2)	−0.014(1)	7

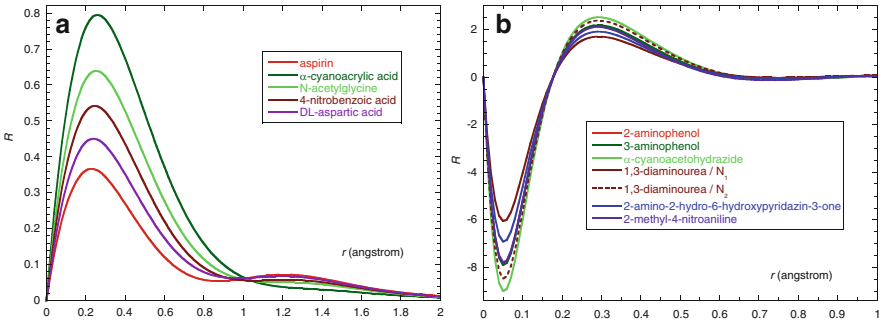


Fig. 5 Selected radial function for chemically similar stockholder atoms before averaging: (**a**) H_R_{11}_[COOH], (**b**) N_R_{11}_[NH_2]

quadrupolar, that of the H-atoms is of dipolar, while that of the N-atom is of octupolar symmetry.

Figure 5 depicts the *H_R$_{11}$_[COOH]* and *N_R$_{11}$_[NH_2]* functions for which the largest discrepancies are found before averaging. The peak-height differences observed for some of the SPA-RDFs for "equivalent" atoms suggests that the stockholder decomposition is quite sensitive to fine details in the chemical environment. The consistency in the peak locations, on the other hand, might imply that these environmental differences manifest themselves only locally.

In general, the monopole SPA-RDFs show a high degree of chemical transferability. Even for the H-atoms, only relatively small variations of the *H_R_{00}[X–H]* functions with the neighboring atom type (*X*) are observed, but only near to the proton's position. Figure 6 displays SPA-RDFs averaged over chemical equivalents and corresponding to the dominating deformation densities for each atom. The averaged *H_R_{10}[X–H]* functions associated with the bond-directed dipolar deformation of the H-atoms have quite similar shape for all *X–H* bonds (Fig. 6a). The location of the maxima (~0.25 Å) seems to be independent of the bonding situation, but the peak-height scatters in a wide range (0.6–1.6) and correlates with the

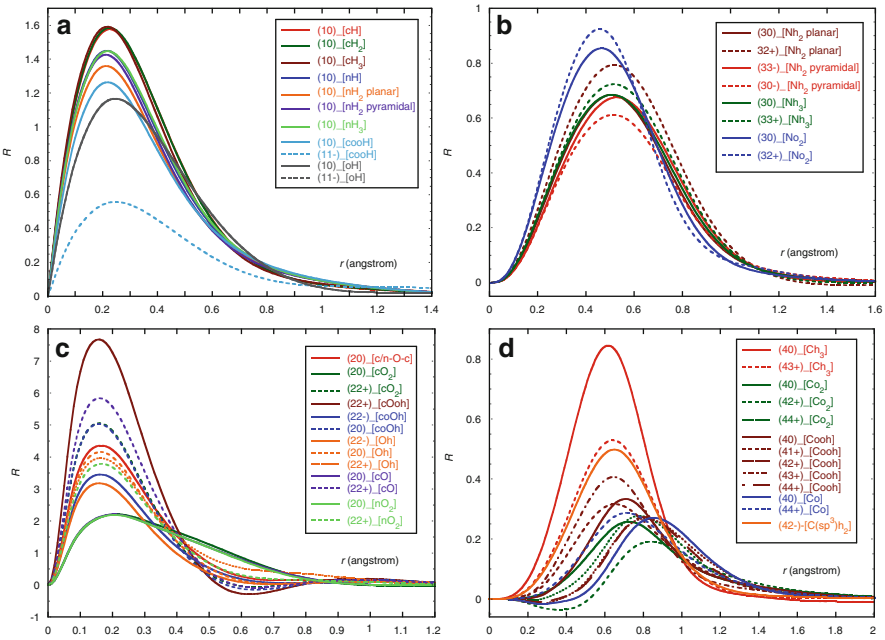


Fig. 6 Selected radial functions averaged over equivalent stockholder atoms: (**a**) H_R_{1m}, (**b**) N_R_{3m}, (**c**) O_R_{2m}, (**d**) C_R_{4m}

electronegativity of X. In other words, $H_R_{10}[C–H]$ is sharper than $H_R_{10}[O–H]$. The higher-order SPA-RDFs seem to be more sensitive to the chemical environment and thus are less transferable than the monopoles and dipoles.

The best agreement between various types of N-atoms is found for the octupole functions (Fig. 6b); the height of the maxima is very consistent, but their location is slightly shifted closer to the N-nucleus in the nitro group. However, some higher-order SPA-RDFs can be markedly different. For example, the $N_R_{10}[NH_3^+]$ shows a behavior that is opposite to that of $N_R_{10}[NH_2]$; the maxima and minima are interchanged, while the locations of the extrema are preserved. The quadrupole functions in the two types of amino groups (planar and pyramidal) are different not only in terms of the height-maxima. For the nitro group, they depart considerably from those in the two amino groups. The behavior of hexadecapole functions is similar to that of quadrupoles.

All but three quadrupolar functions for the O-atoms (Fig. 6c) show very similar radial behavior and peak almost at the same location, but may differ in the height-maximum: $O_R_{20}[COO^-]$, $O_R_{20}[C{=}O]$ and $O_R_{20}[NO_2]$. The main maximum of the octupole SPA-RDFs is observed at the same distance from the nuclei ($r \sim 0.5$ Å) for all types of O-atoms. Several of the octupole functions also show a second maximum, sometimes separated by a node from the first one, while the majority of the functions exhibit monotonic decay. A similar picture is observed for haxadecapoles; most of these functions have only one maximum, change sign as

approaching to zero. However, all three $O_R_{42}[E]$ functions ($E = CO_2$, $COOH$, and $C{=}O$) exhibit a well-defined second maximum at $r \sim 1.2$ Å.

For the C-atoms (Fig. 6d), the differences include second minima in $C_R_{11}[C{=}O]$ and $C_R_{10}[COO^-]$. The quadrupole functions split into two groups: one with a maximum close to 0.25 Å, and another with a maximum at around 0.5 Å. As in the case of N-atoms, the octupolar SPA-RDFs show maxima at $r \sim 0.5$–0.6 Å, but unlike for the O-atom, no nodes are observed. The hexadecapolar functions are closer in shape to those found for the N than for the O-atoms.

9 Stockholder Pseudoatom Databank

The fairly good chemical transferability of the SPA-RDFs suggests that it is possible to design radial density basis sets for different type of atoms in a high variety of bonding situations. Such a library appears to be of foremost importance for experimental charge density research if data quality continues to improve at the current rate. Similar efforts using the conventional model have shown that application of HC-PA databanks can be beneficial for routine small- and macro-molecular crystallography [67, 68] and even for molecular modeling [69]. Our approach to SPA database building has many new elements. The density ($\rho^X[E]$) of an X-type SPA in a given bonding environment (E) is approximated by a combination of a transferable part ($\bar{\rho}^X$) and a correction to it ($\delta^X[E]$):

$$\rho^X[E] = \bar{\rho}^X + \delta^X[E] = \frac{1}{N_E}\sum_E \rho_E^X + \sum_{lm} Q_{lm}^X R_l^X d_{lm} \tag{19}$$

The transferable term is the average SPA taken over a set of equivalents ($\{\rho_E^X\}_{N_E}$) derived from a large number of reference molecules each containing the X-type embedded in slightly different bonding environments. The RDFs of the nontransferable ($\delta^X[E]$) part are derived via minimizing the RMS composed of residual EDs not accounted for by the average term:

$$\varepsilon^X = \sum_E \int \left\{\left(\rho_E^X - \bar{\rho}^X\right) - \delta^X[E]\right\}^2 \mathrm{d}\mathbf{r} \tag{20}$$

The m-independent RDFs of the correction terms ($\delta^X[E]$) are expanded over an auxiliary Slater basis, which is derived by the principal component analysis of the set of equivalent SPAs $\{\rho_E^X\}_{N_E}$.

The database building is a five-step procedure:

1. Selection of reference molecules containing the atoms to be included in the library. The current version was built from about 1,800 molecules composed of second period (B–F) and some heavier elements, such as P, S, and Cl. Accurate structures were taken from the CSD, as described above.

2. Identification of and search for equivalent atoms. An automatic protocol scans the reference structures and collects equivalent atoms using different structure descriptors and similarity measures. These descriptors include the atomic number, the local connectivity graph and bond-order/aromaticity indices. Based on the statistical agreement of similarity measures, a decision is made for each atom whether it belongs to an existing group or it is to be considered as a new type. For each atom, grouped in such a way, the descriptors are stored together with a code referring to the molecule the atom originates from.
3. Ab initio electronic structure calculations on reference molecules containing atoms picked up by the above protocol. The molecular densities, used to construct the current databank, were generated at the B3LYP/aug-cc-pVTZ level.
4. Calculation of the RDFs. For each type of SPAs, the RDFs are calculated from the molecular wave functions on a fine radial grid. This step also performs a statistical analysis and the fit to obtain the analytic SPA-RDFs of the average and correction densities.
5. Building the ED of the target molecule of a known structure. Input atomic positions and specific criteria for matching similarity indices are used by the builder to construct both the average density ($\bar{\rho}^X$) and the RDFs of the first-order correction. The total charge is carried by the average term whose monopole populations are scaled so as to re-establish the electro-neutrality.

The applicability of the SPA databank is being explored using both simulated and experimental structure factors for molecules of 60–70 atoms, which can still be challenging targets of X-ray charge density analyses. According to (19), the library-based SPA protocol keeps the transferable density fixed and completes the data analysis in terms of the nontransferable part. Since this option is not yet implemented into XD, our preliminary studies used only the averaged SPA component. Results of the interpretation of theoretical (B3LYP/cc-pVTZ) and experimental data (100K, synchrotron radiation, at 1.25 Å^{-1} resolution [70]) of the *Terbogrel* ($C_{23}H_{27}N_5O_2$) molecule (Fig. 7d) are presented here as an example.

The structure factors calculated with the average SPAs drawn from the databank are in an excellent agreement ($R = 0.52\%$) with those generated from the wave function. Figure 7a displays the direct-space residual density for the guanidine fragment, which exactly maps features unaccounted for by the zero-order (average) database model. This is the section of the residual ED where the worst agreement between the "exact" and the library densities was found, as seen by the pronounced residual contours around the N and C-atoms of the –N–C≡N group. The refinement of multipole populations of the average density lowers the *R*-value to 0.38% and removes the highest residual contour lines (Fig. 7b). No bonding features higher than 0.05 eÅ^{-3} are left. While the accuracy of the fitted average density is quite satisfactory, the remaining density features can be almost completely removed by adjusting the multipole populations of $\delta^X[E]$ to the residual structure factors $(F(\rho_\Psi) - F(\bar{\rho}))$. The ADP refinement of the average SPAs against the experimental

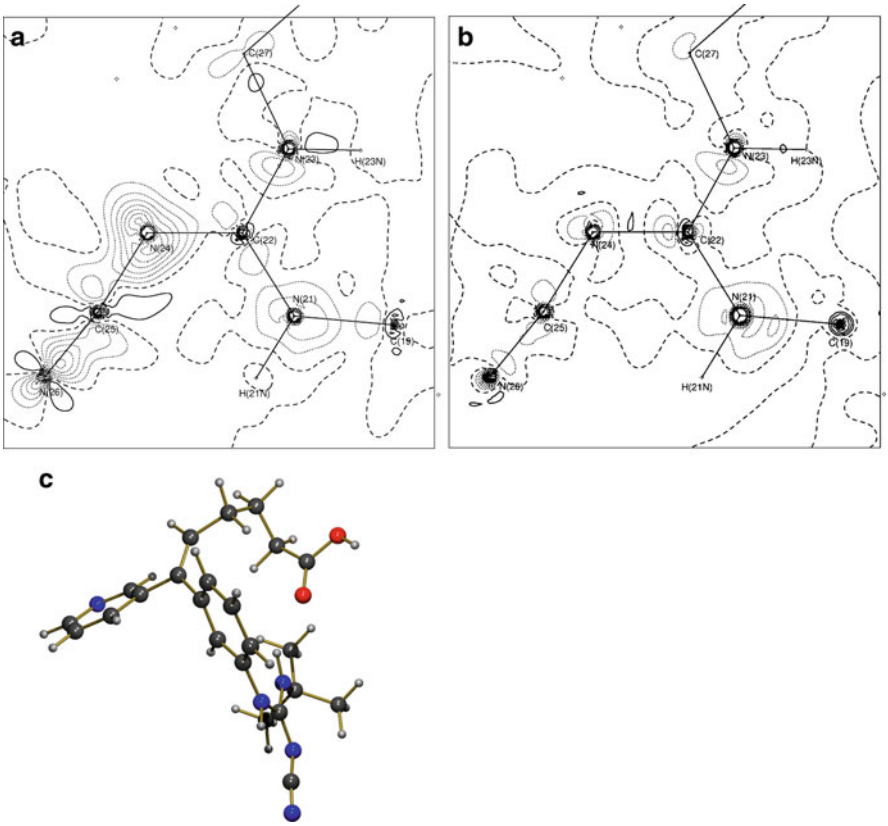


Fig. 7 Direct-space difference density maps for *Terbogrel* (**c**) in the plane of the guanidine fragment. (**a**) $\rho_\Psi - \rho_{\mathrm{AVE}}$, where ρ_{AVE} is the average density built from the SPA data bank, (**b**) the same map obtained after refinement of the populations of the average databank density. The theoretical density was calculated at the B3LYP/cc-pVTZ level. *Contour lines* are drawn at 0.05 eÅ^{-3}. *Solid*, *dashed*, and *dotted lines* are positive, zero, and negative, respectively

data yields an *R*-value of 2.02%, which decreases to 1.65% after the refinement of the SPA populations.

10 Summary

As demonstrated by the *glycine* simulation, the unrestricted SPA ($L = 4$) model provides an adequate representation of the MO density and the corresponding error-free static structure factors. The study also implies that the application of bonded-atom RDFs is necessary to reach the accuracy commonly claimed for experimental pseudoatom densities. However, the accuracy reachable for the local topological figures seems to be severely limited by the level of expansion. This could partially

be overcome with the inclusion of higher-order deformation functions in the model. Stockholder pseudoatoms with *m*-independent RDFs cannot reconstruct the target density to the same accuracy, yet fortuitous error cancellations can occasionally yield better BCP estimates than those obtained via the unconstrained model.

The direct-space projection technique is an efficient validation tool applicable to methodological developments in X-ray charge density research. Analyses of synthetic (theoretical) structure factor data of increasing complexity can help optimize the refinement protocol and minimize the bias in the fitted density. An important problem to be studied is the deconvolution of thermal smearing effects from structure factors. The stockholder decomposition, yielding a one-center representation of the target density, provides a relatively easy route to thermal smearing. It becomes also possible to incorporate anharmonicity into the simulated structure factor data and to analyze its effects on the fitted model density. An equally important issue to be addressed is the extent to which crystal-field effects are detectable via the pseudoatom model. This can be assessed in detail by interpreting structure factors from periodic calculations in terms of the SPA representation of the wave-function-based density of the isolated molecule.

Theoretical pseudoatoms allow for probing experimental structure factors of small-molecule crystals against a known density of the isolated molecule and thus help identify systematic errors and design refinement protocols to account for them. The improved least-squares convergence and stability achieved using bonded-atom RDFs is accompanied by a considerable decrease in correlation between the parameter estimates, making thus possible to reduce, or even avoid, parameter restrictions routinely invoked to condition the fit. Chemical constraints, for example, are commonly enforced in the course of standard pseudoatom-based modeling, not only to decrease the number of variables to be refined but also to eliminate parameter indeterminacies. The latter issue is critical for non-centrosymmetric structures [71], in which case the indeterminacy of the odd-order populations usually results in an ill-conditioned least-squares matrix. In general, it is anticipated that the interpretation of high-quality data with the upgraded multipole formalism has the potential of accessing information that remains hidden for the standard model due to its limited resolving power.

The two small-molecule examples presented here and other studies with similar outcomes show, however, that even a medium-level ab initio density of the isolated system can well account for a medium-quality and -resolution experimental data. One might arrive at the conclusion that important signals of crystal-field effects can stay below the experimental error level and/or "washed out" by thermal motion and resolution effects. These observations suggest that the influence of weak intermolecular interactions on the intramolecular ED should be interpreted with caution.

Theoretical pseudoatoms can provide a good starting model also for treating polarized neutron data to derive the spin (magnetization) density of molecular crystals. Such an analysis should use two sets of RDFs; one projected from the majority (α) and one from the minority (β) spin densities calculated from the corresponding spin-orbitals at the unrestricted level of theory. This spin-separated pseudoatom formalism, yielding models for the charge and spin densities consistent

with the wave function, can be especially advantageous for a joint treatment of X-ray and polarized neutron data [72]. The databank of theoretical pseudoatoms can open new perspectives for routine X-ray charge density analyses, especially of larger molecular systems, provided technical advancements further improve data quality.

Acknowledgment This work was supported by the Deutsche Forschungsgemeinschaft (Priority program 1178: "Experimental electron density as the key to understand chemical bonding").

References

1. Coppens P (1997) X-ray charge densities and chemical bonding. Oxford University Press, Oxford
2. Tsirelson VG, Ozerov RP (1996) Electron density and bonding in crystals. Institute of Physics Publishing, Bristol
3. Stewart RF, Feil D (1980) A theoretical study of elastic X-ray scattering. Acta Cryst A36:503–509
4. Stewart RF (1977) Vibrational averaging of X-ray-scattering intensities. Isr J Chem 16:137–143
5. Stewart RF (1977) One-electron density functions and many-centered finite multipole expansions. Isr J Chem 16:124–131
6. Hansen NK, Coppens P (1978) Testing aspherical atom refinements on small molecule data sets. Acta Cryst A34:909–921
7. Koritsanszky TS, Coppens P (2001) Chemical applications of X-ray charge density analysis. Chem Rev 101:1583–1628
8. Bader RFW (1990) Atoms in molecules: a quantum theory. Oxford Science Publications, Oxford
9. Abramov YuA (1997) On the possibility of kinetic energy density evaluation from the experimental electron-density distribution. Acta Cryst A53:264–272
10. Espinosa E, Mollins E, Lecomte C (1998) Hydrogen bond strengths by topological analyses of experimentally observed electron densities. Chem Phys Lett 285:170–173
11. Tsirelson VG (2001) The mapping of electronic energy distributions using experimental electron density. Acta Cryst B58:632–639
12. Tsirelson VG, Stash AI (2004) On functions and quantities derived from the experimental electron density. Acta Cryst A60:418–426
13. Tsirelson VG, Stash AI, Potemkin VA, Rykounov AA, Shutalev AD, Zhurova EA, Zhurov VV, Pinkerton AA, Gurskaya GV, Zavodnik VE (2006) Molecular and crystal properties of ethyl 4, 6-dimethyl-2-thioxo-1, 2, 3, 4-tetrahydropyrimidine-5-carboxylate from experimental and theoretical electron densities. Acta Cryst B62:676–688
14. Zhurova EA, Stash AI, Tsirelson VG, Zhurov VV, Bartashevich EV, Potemkin VA, Pinkerton AA (2006) Atoms-in-molecules study of intra- and intermolecular bonding in the pentaerythritol tetranitrate crystal. J Am Chem Soc 128:14728–14734
15. Zhurova EA, Zhurov VV, Pinkerton AA (2007) Structure and bonding in β-HMX-characterization of a trans-annular N···N interaction. J Am Chem Soc 129:13887–13893
16. Jelsch C, Pichon-Pesme V, Lecomte C, Aubry A (1998) Transferability of multipole charge-density parameters: application to very high resolution oligopeptide and protein structures. Acta Cryst D54:1306–1318
17. Zarychta B, Pichon-Pesme V, Guillot B, Lecomte C, Jelsch C (2007) On the application of an experimental multipolar pseudo-atom library for accurate refinement of small-molecule and protein crystal structures. Acta Cryst A63:108–125

18. Abramov YuA, Okamura FP (1997) A topological analysis of charge densities in diamond, silicon and germanium crystals. Acta Cryst A53:187–198
19. Tsirelson VG, Stash AI (2002) Topological analysis of electrostatic potential in $SrTiO_3$. Acta Cryst B58:567–575
20. Ivanov YuV, Belokoneva EL, Protas J, Hansen NK, Tsirelson VG (1998) Multipole analysis of the electron density in topaz using X-ray diffraction data. Acta Cryst B54:774–781
21. Lippmann T, Blaha P, Andersen NH, Poulsen HF, Wolf T, Schneider JR, Schwarz K-H (2003) Charge-density analysis of $YBa_2Cu_3O_{6.98}$. Comparison of theoretical and experimental results. Acta Cryst A59:437–451
22. Kirfel A, Krane H-G, Blaha P, Schwarz K, Lippmann T (2001) Electron-density distribution in stishovite, SiO_2: a new high-energy synchrotron-radiation study. Acta Cryst A57: 663–677
23. Coppens P, Iversen B, Larsen FK (2005) The use of synchrotron radiation in X-ray charge density analysis of coordination complexes. Coord Chem Rev 249:179–195
24. Macchi P (2009) Electron density distribution in organometallic materials. Chimia 63:29–34
25. Macchi P, Sironi A (2003) Chemical bonding in transition metal carbonyl clusters: complementary analysis of theoretical and experimental electron densities. Coord Chem Rev 238:383–412
26. Fertig HA, Kohn W (2000) Symmetry of atomic electron density in Hartree, Hartree-Fock and density-functional theories. Phys Rev A62:052511-10
27. Lennard-Jones JE (1929) The electronic structure of some diatomic molecules. Trans Faraday Soc 25:668–686
28. Stewart RF, Bentley J, Goodman B (1975) Generalized X-ray scattering factors in diatomic molecules. J Chem Phys 63:3786–3793
29. Clementi E, Roetti C (1974) At Data Nucl Data Tables 14:177–576
30. Clementi E, Raimondi DL (1963) Atomic screening constants from SCF functions. J Chem Phys 38:2686–2689
31. Autschbach J, Schwarz WHE (2000) Where do the forces in molecules come from? A density functional study of N_2 and HCl. J Phys Chem A104:6039–6046
32. Brown AS, Spackman MA (1991) A model study of κ-refinement procedure for fitting valence electron densities. Acta Cryst A47:21–29
33. Pèrés N, Boukhris A, Souhassou M, Gavoille G, Lecomte C (1999) Electron density in ammonium dihydrogen phosphate: non-uniqueness of the multipolar model in simple inorganic structures. Acta Cryst A55:1038–1048
34. Pillet S, Souhassou M, Lecomte C, Schwarz K, Blaha P, Rerat M, Lichanot A, Roversi P (2001) Recovering experimental and theoretical electron densities in corundum using the multipolar model: IUCr multipole refinement project. Acta Cryst A57:290–303
35. Bader RFW, Essèn H (1984) Characterization of atomic interactions. J Chem Phys 80:1943–1960
36. Flaig R, Koritsanszky T, Dittrich B, Wagner A, Luger P (2002) Intra- and intermolecular topological properties of amino acid. A comparative study of experimental and theoretical results. J Am Chem Soc 124:3407–3417
37. Howard ST, Hursthouse MB, Lehmann CW (1995) Experimental and theoretical determination of electronic properties in L-dopa. Acta Cryst B51:328–337
38. Spackman MA, Byrom PG (1997) Retrieval of structure-factor phases in non-centrosymmetric space group. Model studies using multipole refinement. Acta Cryst B53:553–564
39. Spackman MA, Byrom PG, Alfredsson M, Hermansson K (1999) Influence of intermolecular interactions on multipole refined electron densities. Acta Cryst A55:30–47
40. Volkov A, Coppens P (2001) Critical examination of the radial functions in the Hansen-Coppens multipole model through topological analysis of primary- and refined-theoretical densities. Acta Cryst A57:395–405
41. Bytheway I, Chandler SG, Figgis BN (2002) Can a multipole analysis faithfully reproduce topological descriptors of a total charge density? Acta Cryst A58:451–459

42. Volkov A, Abramov YA, Coppens P (2001) Density optimized radial exponents for X-ray charge density refinement from ab initio calculations. Acta Cryst A57:272–282
43. Dominiak PM, Coppens P (2006) Finding optimal radial-function parameters for S atoms in the Hansen-Coppens multipole model through refinement of theoretical densities. Acta Cryst A62:224–227
44. Volkov A, Macchi P, Farrugia LJ, Gatti C, Mallinson P, Richter T, Koritsanszky T (2006) Program manual, XD2006 – A computer program package for multipole refinement, topological analysis of charge densities and evaluation of intermolecular energies from experimental and theoretical structurefactors. User's manual. http://xd.chem.buffalo.edu/docs/xdmanual.pdf
45. Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46:618–622
46. Dunning TH Jr (1989) Gaussian basis sets for use in correlated molecular calculations. I. the atoms boron through neon and hydrogen. J Chem Phys 90:1007–1023
47. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03W, Revision C.02. Gaussian, Inc., Wallingford CT
48. Lee C-R, Tang T-H, Chen L, Wang C-C, Wang Yu (2004) Topological analysis of charge density in heptasulfur imide (S_7NH) from isolated molecule to solid. J Phys Chem Solids 65:1957–1966
49. Koritsanszky T, Volkov A (2004) Density radial functions for bonded atoms. Chem Phys Lett 385:431–434
50. Hirshfeld FL (1977) Spatial partitioning of charge density. Theor Chim Acta 44:129–132
51. Nalewajski RF, Parr PG (2000) Information theory, atoms in molecules and molecular similarity. Proc Natl Acad Sci USA 97:8879–8882
52. Destro R, Roversi P, Barzaghi M, Marsh RE (2000) Experimental charge density of α-glycine at 23 K. J Phys Chem A104:1047–1054
53. te Velde B, Bickelhaupt FM, van Gisbergen SJA, Fonseca Guerra C, Baerends EJ, Snijders JG, Ziegler TJ (2001) Chemistry with ADF. J Comput Chem 22:931–967
54. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B37:785–789
55. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic-behavior. Phys Rev A38:3098–100
56. van Lenthe EV, Baerends EJ (2003) Optimized Slater-type basis sets for the elements 1-118. J Comput Chem 24:1142–1156
57. Sùrensen HO, Stewart RF, McIntyre GJ, Larsen S (2003) Simultaneous variation of multipole parameters and Gram-Charlier coefficients in a charge-density study of tetrafluorotere-phthalonitrile based on X-ray and neutron data. Acta Cryst A59:540–550
58. Koritsanszky T, Buschmann J, Luger P (1996) Topological analysis of experimental electron densities (1): the different C-C bonds in Bullvalene. J Phys Chem 100:10547–10552
59. Dittrich B, McKinnon JJ, Warren JE (2008) Improvement of anisotropic displacement parameters from invariom-model refinements for three L-hydroxylysine structures. Acta Cryst B64:750–759
60. Meindl K, Henn J (2008) Foundations of residual-density analysis. Acta Cryst A64:404–418

61. Volkov A, Li X, Koritsanszky T, Coppens P (2004) Ab initio quality electrostatic atomic and molecular properties from a transferable theoretical pseudoatom databank: Comparison of electrostatic moments, topological properties, and interaction energies with theoretical and force-field results. J Phys Chem 108:4283–4300
62. Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Cryst B58:380–388
63. Allen FH (1986) A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron diffraction. Acta Cryst B42:515–522
64. Lebedev VI, Laikov DN (1999) A quadrature formula for the sphere of the 131st algebraic order of accuracy. Dokl Math 59:477–488
65. Perez-Jorda JM, San-Fabian E, Moscardo F (1992) A simple, reliable and efficient scheme for automatic numerical integration. Comput Phys Commun 70:271–284
66. Treutler O, Ahlrichs R (1995) Efficient molecular numerical integration schemes. J Chem Phys 102:346–354
67. Pichon-Pesme V, Lecomte C, Lachekar H (1995) On building a data bank of transferable experimental electron density parameters: applications to polypeptides. J Phys Chem 99:6242–6250
68. Dittrich B, Koritsanszky T, Luger P (2004) A Simple approach to molecular densities with invarioms. Angew Chem Int Ed 43:2718–2721
69. Dominiak PM, Volkov A, Dominiak AP, Jarzembska KN, Coppens P (2009) Combining crystallographic information and an aspherical-atom data bank in the evaluation of the electrostatic interaction energy in an enzyme-substrate complex: influenza neuraminidase inhibition. Acta Cryst D65:485–499
70. Flaig R, Koritsanszky T, Soyka R, Häming L, Luger P (2001) High-resolution X-ray crystallography in drug design: electronic properties of an antithrombotic agent. Angew Chem Int Ed 40(2):355–359
71. Haouzi AE, Hansen NK, Hènass CLe, Protas J (1996) The phase problem in the analysis of X-ray diffraction data in terms of electron-density distributions. Acta Cryst A52:291–301
72. Becker P, Coppens P (1985) About the simultaneous interpretation of charge and spin density data. Acta Cryst A41:177–182

Struct Bond (2012) 147: 27–46
DOI: 10.1007/430_2012_78
© Springer-Verlag Berlin Heidelberg 2012
Published online: 27 March 2012

Reliable Measurements of Dipole Moments from Single-Crystal Diffraction Data and Assessment of an In-Crystal Enhancement

B. Dittrich and D. Jayatilaka

Abstract Using seven examples of high-quality data sets of amino acids it is shown that accurate molecular dipole moments can be obtained from experimental diffraction data. Recommendations for practical modeling choices are given when using the Hansen/Coppens multipole model. Multipole-model results, including those from invariom refinement, are found to be less accurate than results from a basis-set description. The question whether a molecular dipole-moment enhancement in the solid state is fact or artifact is studied by a number of techniques: A theoretical molecule embedded in a cluster of point-charges gives a substantial enhancement, in agreement with Hirshfeld atom refinement with point charges and dipoles. The experimental techniques, multipole refinement and wavefunction fitting, lead to smaller dipole-moment enhancements than the theoretical predictions.

Keywords Single-crystal X-ray diffraction • molecular dipole moment • dipole-moment enhancement • multipole model • wavefunction fitting • hirshfeld-atom refinement

Contents

1 Introduction 28
2 Experimental Datasets 30
2.1 Experimental Challenge: Hydrogen Scattering 30
2.2 Experimental Challenge: Data Resolution 31

B. Dittrich (✉)
Institut für Anorganische Chemie der Universität Göttingen, Tammannstr. 4, 37077 Göttingen, Germany
e-mail: bdittri@gwdg.de

D. Jayatilaka
Chemistry, M313, School of Biomedical, Biomolecular and Chemical Science, University of Western Australia, WA 6009 Crawley, Australia

3 On the Ability of the Multipole Model to Reproduce Theoretical Dipole Moments 32
4 Dipole-Moment Enhancements from Theory 34
4.1 Dipole-Moment Enhancements from Simple Theoretical Cluster Calculations 35
4.2 Theoretical Estimate of Dipole-Moment Enhancements with Cluster Charges and Dipoles 37
5 Dipole-Moment Enhancements by Combining Theory and Experiment 37
5.1 Molecular Dipole Moments and Their Enhancement in the Solid State from Experimental Multipole Refinement and Invariom Refinement 38
5.2 Molecular Dipole Moments and Their Enhancement from Hirshfeld-Atom Refinement and Wavefunction Fitting 40
6 Discussion: Agreement Between Experimental and Theoretical Results 41
7 Conclusion 42
References 43

1 Introduction

Intermolecular forces are of great interest in chemistry and physics. The classical electrostatic interaction energy between two species can be expanded in a multipole series. Its most important term (for neutral species) is the dipole moment [1]. The dipole of a system is of fundamental and continuing interest.

When non-spherical scattering models were introduced in the late-1960s [2–4] and optimized throughout the 1970s [5–7] it became possible to obtain dipole and higher multipole moments from accurate single-crystal X-ray diffraction data. The basic characteristic common to these different non-spherical scattering models is that they provide an analytical description of the electron density distribution $\rho(\mathbf{r})$ (EDD) in terms of products of atom-centred radial and spherical harmonic angular functions. Only the populations of the latter angular functions (and possibly a radial scaling parameter κ) are adjusted (via a least-squares procedure) to reproduce the intensities of the diffraction experiment. The Hansen/Coppens approach [7] has proven to be successful throughout the last decades in that it has enabled experimental characterization of solid state electronic structure and bonding.

Lately, the multipole model [5, 7] has undergone significant development and a change in philosophy. Instead of the multipole parameters being refined from the X-ray data, they can alternatively be predicted by fitting to theoretical data obtained from quantum mechanical calculations [8]. Not only the multipole parameters but also H-atom vibration parameters (the atomic displacement parameters or ADPs) can additionally be derived from theoretical calculations or other external sources of information like neutron diffraction [9–11]. Programs and schemes have been developed to transfer electron-density parameters from atoms in smaller molecules into larger molecules where the chemical environment is similar. Thus, the traditional role of the experimental measurements determining bonding density has been depreciated in favour of an emphasis on accurate geometric parameters, especially for larger molecules. The significance of these developments on dipole-moment determination from X-ray diffraction data requires substantial characterization.

Even more recently, sophisticated quantum mechanical methods have been developed to refine the geometric and electronic parameters of the crystal structure. For example, it is now possible to refine geometric parameters by using non-spherical scattering models based on quantum mechanical calculations [12, 13] (so-called Hirshfeld atom refinement). Remarkably, this leads to accurate X—H bond distances in excellent agreement to distances from Neutron diffraction. It is also possible to combine quantum mechanical methods directly with the least-squares refinement of the electronic structure parameters describing the electron density (X-ray constrained wavefunction methods) [8, 14–20]. Again, the impact of these methods on dipole-moment determination from X-ray diffraction data merits further study.

Earlier studies aimed at obtaining an experimental dipole moment from diffraction data by refinement of multipole parameters have been comprehensively reviewed in the past [21, 22]. A mathematical definition of the dipole moment and detailed background information can also be found in these review articles. The common consensus is that obtaining reliable dipole moments is a "challenging" undertaking but certainly worthwhile, because the diffraction experiments "are unrivalled in their potential to provide this information in such detail" [22]. This latter comment refers to the fact that, unlike in many other experiments, all the components of the dipole moment are determined from an X-ray diffraction study. Further, dipole moments of molecular fragments can be obtained.

Nevertheless, dipole-moment determinations from multipole refinement frequently remain unreliable, with enhancements in the dipole moment in excess of $\geq$100% having been reported. Several reasons for this have been clearly enunciated [22] including the fact that the definition of a dipole moment in a crystal from a periodic charge density requires a well-defined partitioning of a molecule in a crystal [23].[1] Further limitations include data quality, especially for data pertaining to non-centrosymmetric crystals where phases are less well determined [25, 26], and – what is of interest in this paper – limitations in the modelling process.

In this article we seek to characterize the situations in which an accurate dipole moment can be determined from X-ray diffraction data using the multipole model, Hirshfeld-atom refinement and X-ray constrained wavefunctions. Several questions are addressed:

- *What are the expected accuracies for dipole moment magnitudes? Are there possible pitfalls?*
  We investigate this question by fitting the multipole model to static structure factors for 22 small organic molecules.
- *What are the accuracies for dipole moments determined from multipole-model scattering-factor databases?*
  Structure refinements with scattering-factor databases like the invariom database [27] offer rapid access to dipole moments, and it is important to quantify their performance with respect to dipole-moment evaluation. This is achieved by comparison with experimental results (from refined multipoles) in Table 5.

[1] It must be noted that definitions can be made for the unit-cell polarization, which are independent of the charge density and hence are well defined for periodic systems [24].

Table 1 Crystallographic details of the structures studied. The radiation (Rad.) used and the resolution (Res.) reached (in $\sin\theta/\lambda$, i.e. in Å^{-1}) are given

Structure	Spacegr.	Z, Z′	Temp.	Rad.	Res.	Ref.
L-Alanine	$P2_12_12_1$	4,1	23 K	Mo Kα	1.08	[30]
L-Cysteine	$P2_12_12_1$	4,1	30 K	Mo Kα	0.72	[31]
L-Glutamine	$P2_12_12_1$	4,1	100 K	Mo Kα	1.08	[32]
D,L-Serine	$P2_1/a$	4,1	20 K	Mo Kα	1.19	[33]
L-Threonine	$P2_12_12_1$	8,1	19 K	Ag Kα	1.35	[34]
D,L-Aspartic Acid	$C2/c$	8,1	20 K	Ag Kα	1.37	[35]
D,L-Histidine	$P2_1/c$	4,1	100 K	Mo Kα	1.22	[28]

- *What are the accuracies for dipole moments determined from experimental data (refined multipoles/wavefunction fitting)? What in-crystal enhancements may be expected when compared to the theoretical prediction (invarioms/single-point calculation)? Can theory provide a benchmark to discern enhancements of dipole moments being "fact or artefact"?*
  We address these issues by investigating the dipole moments for seven amino acids for which X-ray diffraction data were obtained from the original authors or were available in the literature. These are compared to reference values obtained from invariom refinement, from ab initio quantum mechanical calculations for isolated molecules and for molecules in a crystal environment. These latter are obtained from self-consistent crystal-field embedded molecular ab initio quantum mechanical calculations. We have also used the X-ray constrained wavefunction method to produce benchmark dipole moments as an alternative to the multipole model. Finally, variations in the dipole moment due to geometric positions from different refinement models are investigated.

2 Experimental Datasets

The structures of the genetically encoded amino acids have been extensively studied. However, dipole moments from X-ray diffraction have not frequently been reported for these molecules, with the exceptions of D,L-histidine and L-alanine [28, 29]. We have therefore chosen to focus on accurate structure determinations of seven amino acids previously reported in the literature for our study (Table 1). In all cases only one single molecule crystallizes in the asymmetric unit. Molecules chosen are L-alanine [30], L-cysteine [31], L-glutamine [32], D,L-serine [33], L-threonine [34], D,L-aspartic acid [35] and D,L-histidine [28]. High-resolution data were provided by the respective authors or were available electronically. In the case of L-cysteine, high-resolution data were not determined.

2.1 Experimental Challenge: Hydrogen Scattering

Even when carefully modelling the information content of the Bragg data, complications in determining dipole moments are likely to arise due to the X-ray

scattering properties of hydrogen atoms. These properties, comparably weak scattering with limited resolution in reciprocal space, for decades have been known to cause concern regarding the reliability of properties based on least-squares refined parameters of H-atoms [36]. Since H-atoms are often situated at the molecular periphery and often far away from the centre of mass, their influence on the molecular dipole moment can be significant.[2] Accurate X–H bond distances are therefore imperative. Neutron diffraction experiments are the preferred source of accurate X–H bond distances; but results for particular molecules or bonding environments are usually not available due to the considerable additional experimental effort. Favourable developments with the advent of spallation Neutron sources might change this situation in the future.

Technical improvements help to reduce the problem caused by the scattering properties of hydrogen. A recent study used external information from periodic calculations to try to limit the flexibility of the screening parameters [37] for C, N and O. Scattering-factor databases [27, 38, 39] provide even more accurate "hybrid" scattering factors, also for hydrogen atoms. For the theoretical databases [27, 39] these hybrid scattering factors are obtained by combining fixed multipoles from the database – with the order of the expansion $l \geq 1$ – with refined monopole and dipole populations. In that sense hybrid scattering factors for H-atoms can be seen in analogy to constraints or restraints, since they reduce the flexibility of the least-squares refinement model by adding prior chemical information. Furthermore, X–H distances from geometry optimizations can now be used. They are included in the invariom database [27] and can be retrieved with the program INVARIOMTOOL [40]. In Sect. 5.1 we show how hybrid hydrogen scattering factors and fixed X–H bond distances can increase the reliability of the determination of dipole moments from multipole refinements.

2.2 Experimental Challenge: Data Resolution

Apart from a careful treatment of hydrogen scattering, another requirement for the determination of reliable dipole moments from multipole refinement of X-ray diffraction is good quality low-temperature intensity data, preferably extending high into reciprocal space. These are required in order to refine the large number of possible least-squares parameters per atom (three positional, six displacement and up to 25 multipole parameters up to $l_{max} = 4$, not counting radial screening parameters). Low temperature is mandatory, since experimental conditions are more favourable, e.g. regarding the significant reduction of atomic displacements and thermal diffuse scattering [41]. For further (experimental) requirements concerning multipole refinements of X-ray diffraction we refer to [42].

[2] One way to resolve the issue is to choose a sample devoid of hydrogen altogether.

Since high-resolution data were not available in the case of L-cysteine, we use recently introduced methodology [43] to obtain an experimental dipole moment despite limitations in data resolution by including ADPs from a previous invariom refinement [33] in a block-matrix refinement of L-cysteine. This procedure and the low data-collection temperature of 30 K allowed to reach the same accuracy as achieved for the other examples.

Requirements for data resolution are more modest for invariom refinements and when using other databases. Nevertheless, despite the success of the suggested block-matrix refinement procedure for L-cysteine, high-resolution data are certainly preferred or even required for the multipole refinements used in our comparative studies.

3 On the Ability of the Multipole Model to Reproduce Theoretical Dipole Moments

The initial question raised is simple: How well does the multipole model allow to reproduce theoretical dipole moments from a DFT calculation with the B3LYP functional and the comparably extended basis-set D95++(3df,3pd)? In order to answer this question twenty-two molecules exhibiting a dipole moment were chosen (see Table 2 for details). They can be considered representative of organic chemistry with some relevance to biological systems. The test set is neither complete nor exhaustive; e.g. zwitterionic compounds are not part of it. For the amino acids, which are zwitterionic in the solid state, multipole projections of the isolated-molecular dipole moments are given in Sect. 5.1.

Geometries of the test-set molecules were optimized with tight convergence criteria in the program GAUSSIAN [44] followed by a frequency calculation to make sure the global minimum was reached. From the resulting wavefunction, real structure factors for a unit cell with dimensions of 30 Å in space group $P_{\bar{1}}$ were calculated with the program TONTO [45], following a procedure introduced earlier [46]. This way a "projection" of the isolated-molecular density onto the multipole model was achieved. Multipole parameters were then refined using these static theoretical structure factors, "simulating" experimental data. Typical R-Factors from such a refinement are around 0.5% (better when heavier nuclei are present), with residual electron density features less than 0.05 e/Å^3. Better figures of merit cannot be achieved with the standard Hansen/Coppens multipole model, since the core density remains unadjusted unlike in a recent study [47], and since the order l of the multipole expansion is limited to four for the valence region.

In all refinements a consistent refinement strategy was applied. Chemical constraints and local-atomic site symmetry were used where possible. However, it was assured that such choices did not affect the resulting dipole moments when compared to a full refinement of all possible multipoles: differences were found to be negligible. On the other hand, more substantial changes were caused by refining

Table 2 Ability of the multipole model to reproduce dipole moments from theory

Compound	Formula sum	μ_0	μ_1	μ_2	μ_3	μ_4	μ_5
Water	H_2O	1.9	1.5	1.5	2.0	1.6	1.5
Formaldehyde	CH_2O	2.4	2.1	2.0	2.2	2.1	2.0
Methanol	CH_4O	1.7	1.5	1.5	1.8	1.6	1.5
Methaneamine	CH_5N	1.3	1.2	1.2	1.4	1.2	1.2
Formamide	CH_3NO	4.0	3.5	3.3	3.7	3.6	3.4
Formic acid	CH_2O_2	3.9	3.6	3.4	4.0	3.7	3.5
Ethanol	C_2H_6O	1.6	1.2	1.2	1.5	1.3	1.2
Methoxymethane	C_2H_6O	1.3	1.6	1.5	1.7	1.7	1.5
Ethaneamine	C_2H_7N	1.3	1.1	1.1	1.3	1.2	1.1
Acetone	C_3H_6O	3.1	2.8	2.4	3.0	2.9	2.6
Acetamide	C_2H_5NO	3.9	3.6	3.3	3.7	3.7	3.5
Propane-2-ol	C_3H_8O	1.6	1.4	1.3	1.8	1.5	1.3
Acetic acid	$C_2H_4O_2$	4.4	4.2	3.7	4.5	4.3	4.0
2-Methylpropan-2-ol	$C_4H_{10}O$	1.6	1.6	1.4	1.9	1.6	1.5
Methanethiol	CH_4S	1.6	2.1	2.1	2.2	2.1	1.4
Phenol	C_6H_6O	1.3	1.5	1.5	1.8	1.6	1.5
Aniline	C_6H_7N	1.6	2.0	2.1	2.0	2.1	1.9
Ethanethiol	C_2H_6S	1.7	2.6	2.7	2.7	2.4	1.5
Chloromethane	CH_3Cl	2.0	2.2	2.2	2.3	2.3	1.2
Propane-1-thiol	C_3H_8S	1.8	3.0	3.1	3.0	2.7	1.3
Dichloromethane	CH_2Cl_2	1.7	2.2	2.3	2.3	2.3	1.5
Chloroform	$CHCl_3$	1.1	1.7	1.9	1.7	1.7	1.2

Compounds are ordered according to their molecular size. Dipole moments are given directly for the theoretical computation with B3LYP/D95++(3df,3pd) (μ_0), or for different multipole models: μ_1 using κ only, with $l_{max} = 4$ for H, which is the default in the 2006 version of the invariom database [27], μ_2 using an additional shared κ' for $l \geq 1$of all non-H atoms, μ_3 same as μ_1, but limiting $l_{max} = 1$ for H, μ_4, same as μ_1, but limiting data resolution to sin $\theta/\lambda_{max} = 0.8$ Å^{-1}. Very similar values than for μ_1 can be obtained when omitting the shared κ' for carbon atoms and were obtained by keeping the scale factor at 1(μ_5), which improved agreement for S-containing compounds

or not refining the scale factor – which should ideally be unity for theoretical data – and the κ/κ' parameters. Also, the order of the multipole expansion for H-atoms considerably influenced the result obtained (see Table 2). In *experimental* multipole refinements this order l is mostly chosen to be $l \leq 1$ for H, since multipoles with larger l can usually not be refined: As discussed above, correlations and lack of information due to H-atom scattering properties do not allow refinement.

Data resolution can also influence the results. In the currently developed version of the invariom database, simulated data are calculated up to a resolution of sin $\theta/\lambda_{max} = 1.44$ with limiting indices of h, k and l of 50, and cut to a more spherical shell of data to 1.2 Å^{-1} resolution. This procedure was also used here. Results of the different refinements are given in Table 2.

Keeping the level of the multipole expansion at $l \leq 1$ for H-atoms yields a better average agreement for compounds consisting of only C, H, N and O. However, when heavier elements are present, the agreement gets worse and including higher multipoles for hydrogen atoms gives better bond distances in refinement with

experimental data. Furthermore, significant changes in the dipole moments are observed when the resolution is cut to 0.8 $Å^{-1}$ and superior results are often (but not always) obtained in that case.

Improvements with data cut to 0.8 $Å^{-1}$ are probably due to the over-proportional information content of valence electron density in low-order reflections, whereas for heavier elements correlations [43] of the multipole parameters or the frozen core approximation could cause the disagreements seen. It can be observed that dipoles differ most when heavier nuclei like S and Cl are present, and that κ'-parameters are helpful for obtaining a more reliable estimate in such cases. Another factor are Fourier truncations effects, which we are currently investigating. Since the results can deviate by more than 70% (e.g. for chloroform), it is recommended to use fixed κ values from theory in experimental multipole refinements to avoid parameter correlations. Either those fixed κ/κ' values proposed earlier [37, 48] or values obtained from, e.g., the invariom [27] or other databases [38, 39] should be used in our opinion. Fixing the scale factor to unity leads to better agreement with heavier elements present, pointing to the fact that the core density is not well represented by the multipole models' Slater functions in our data generated from Gaussian basis sets. However, fixing some of the "sensitive" model parameters does not generally aid in increasing model flexibility and the ability of the multipole model in reproducing the theoretical dipole moments. It also reduces the characteristic of providing an experimental result.

It is to be expected that the multipole-model dipole moments deviate from the theoretical result, since the density representation used is quite different and more sophisticated in ab initio calculations. In summary one needs to be aware that the classical Hansen/Coppens multipole model cannot fit fine details of the electron density distribution, thereby affecting the dipole moment. Even if an experimental (thermally smeared) electron density might be fitted better than the static structure factors used in this chapter, limitations of the experimental multipole-model approach in accurately reproducing molecular dipole moments become evident.

4 Dipole-Moment Enhancements from Theory

Efforts to theoretically predict changes in the molecular dipole moment when moving from the gas phase to the bulk have initially been challenging, since computations on periodic systems were unfeasible. Nevertheless, elegant predictions based on lattice sums [49, 50] provide good estimates of the effect of crystal packing and hydrogen bonding on molecular electron density [51], despite the approximation of an average uniform electric field, which might be inappropriate for larger molecules and strongly hydrogen-bonded systems. The increase or decrease of the dipole moment has been defined [22] as:

$$\Delta\mu = 100(\mu_{\text{mol. in solid}} - \mu_{\text{single mol.}})/\mu_{\text{single mol.}} \quad (1)$$

Another important step forward in obtaining theoretical solid-state dipole moments was the introduction of Bader's Quantum Theory of Atoms in Molecules (QTAIM) [52], which provides an atomic partitioning scheme for isolated-molecular as well as periodic EDD. One characteristic of Bader's partitioning scheme is that atomic fragments each have a dipole moment. Since the sum of QTAIM fragments and their properties are additive, they reproduce space completely, and a molecular dipole moment can be calculated from the sum of the individual atoms in the gas phase or the bulk. Hence, QTAIM provides an attractive route to accurate dipole moments and their possible enhancements from first principles [53–55]. QTAIM results are not discussed in this study, but are provided, e.g., in [53].

4.1 Dipole-Moment Enhancements from Simple Theoretical Cluster Calculations

The simplest way to obtain dipole-moment enhancements from theory are calculations on molecular clusters which we will now discuss. An obvious approximation made in such an approach is the choice of the distance threshold, for which surrounding whole molecules are included.

For the seven zwitterionic organic molecules studied, a cluster based on a 3–5 Å threshold was used. This corresponds to including all surrounding molecules that are closer than this distance threshold to any atom of the central molecule. Typical cluster sizes, including the examples of the amino acids studied here, are 14–21 molecules. Input files were generated with the program BAERLAUCH [56], and require only atomic positions, a cut-off radius and the space group. To decide which cluster size was required, we geometry optimized the central molecule in the field of surrounding molecules using the ONIOM implementation [57] of quantum mechanics/molecular mechanics (QM/MM) in all seven cases (results not shown here). In case the optimization converged, the cluster size was considered to be sufficient also in single-point cluster calculations. Computational details of the ONIOM procedure for molecular crystals are given in [56].

Calculations with a field of point charges are not expensive to perform, since the environment of the cluster is represented by few additional Gaussian functions. In principle, the method and basis set chosen for the calculations can be as sophisticated and extended as the computer permits. Computational requirements are similar to single-point calculations. Cluster calculations with a field of point charges yield a wavefunction file of an "isolated" molecule. This is in contrast to ONIOM cluster calculations, where the geometry of the central molecule can be optimized, but no isolated-molecule wavefunction file is written in GAUSSIAN [44], since the phase relationship between the different level wavefunctions is undefined. A projection onto the multipole model is technically only possible

Table 3 Total dipole moment (in [D]), individual components and enhancement (in %) for seven amino acids from a simple point charge model with basis-set B3LYP/6-31 G(d,p)

Iteration	x	y	z	Dipole	Enhancement
0	0.3	0.4	−12.4	12.4	–
1	0.7	0.0	−14.4	14.4	16
2	0.7	0.0	−14.7	14.7	19
3	0.7	0.0	−14.8	14.8	19
0	2.5	0.9	11.0	11.3	–
1	3.5	0.5	14.2	14.6	29
2	3.6	0.3	15.0	15.4	36
3	3.7	0.2	15.2	15.7	39
0	10.5	5.0	−4.5	12.4	–
1	13.1	6.1	−4.6	15.2	23
2	13.7	6.2	−4.6	15.7	27
3	13.8	6.3	−4.6	15.9	28
0	−5.4	−11.8	4.8	13.8	–
1	−7.1	−14.9	5.7	17.4	26
2	−7.5	−15.5	5.9	18.2	32
3	−7.6	−15.7	5.9	18.4	33
0	7.7	−4.1	−6.5	10.9	–
1	9.8	−4.8	−8.5	13.8	27
2	10.2	−4.9	−9.2	14.6	34
3	10.3	−4.9	−9.2	14.7	35
0	−4.2	−3.8	−10.0	11.5	–
1	−4.8	−4.9	−12.1	13.9	21
2	−4.9	−5.2	−12.6	14.5	26
3	−4.9	−5.2	−12.7	14.6	27
0	14.3	−0.3	−6.6	15.8	–
1	19.9	1.0	−10.0	22.3	41
2	20.3	1.1	−10.4	22.8	44
3	20.4	1.1	−10.5	22.9	45

Iteration 0 refers to the single molecule only, whereas iteration 1,2 and 3 refer to a calculation, where the atomic point charges from the previous iteration surround the central molecule

when a wavefunction file exists. We therefore did not pursue the ONIOM procedure to obtain dipole-moment enhancements in this study.

The main interest for performing single-point calculations in a cluster of point charges was to get a simple estimate of dipole moments in a cluster, thereby presenting a simple model of a crystal. Therefore, we also limited the size of the basis set to 6-31 G(d,p) in the calculations reported in Table 3. The DFT functional used was B3LYP.

The result from these simple calculations is that substantial dipole-moment enhancements can be observed for all seven amino acids studied. Hence, one could expect them to occur frequently when a molecule becomes part of a crystal lattice. However, these results depend on the approximation of a finite inhomogenous field around the molecule and do not include any experimental information except for the molecular geometry. This result will therefore be verified by experiment and by more sophisticated methodology comparing experiment and theory in Sect. 5.1.

4.2 *Theoretical Estimate of Dipole-Moment Enhancements with Cluster Charges and Dipoles*

A better model for a crystal is accomplished when atomic point charges are complemented by molecular dipole moments in generating the field around a molecule. Like in Sect. 4.1 the purpose is to assess a possible dipole-moment enhancement from a cluster calculation. Apart from including dipole moments of surrounding molecules and from using the program TONTO [45, 46] rather than GAUSSIAN [44], the procedure is analogous. Coordinates after invarioms refinement were chosen as a suitable starting geometry. We note in passing that the HF dipole moment (Table 4) can directly be compared to the invarioms and the DFT single-point dipole moment reported in Table 5. Invarioms aim to reproduce the theoretical values from electron-density fragments. With the basis-set electron density model available in TONTO we can also confirm the well known fact that the Hartree–Fock theory overestimates the dipole moment when compared to calculations that include electron correlation [58]. However, results in Table 4 show that the Hartree–Fock result is a valid estimate and even underestimates the relative in-crystal enhancement seen for DFT. In perspective, molecules studied here exhibit similar enhancement $\Delta\mu$ in the bulk as seen for the point-charge model reported above. Surrounding molecules within a radius of 8 Å were taken into account.

5 Dipole-Moment Enhancements by Combining Theory and Experiment

Experimental determinations of dipole moments usually only provide the value in the solid state. Dipole-moment *enhancements* from experiment can only be obtained by comparing the dipole moment in the solid state with a single-molecule

Table 4 Dipole moments D in Debye from a Hartree–Fock and a DFT calculation on isolated molecules as well as their counterparts in the bulk modelled by a 8 Å cluster of point charges and dipoles

Structure	Basis	HF	HF_{bulk}	$\Delta\mu$/[%]	DFT	DFT_{bulk}	$\Delta\mu$/[%]
L-Alanine	DZP	12.6	17.1	+36	11.1	16.3	+47
	cc-pVTZ	12.2	17.3	+42	10.6	16.6	+57
L-Cysteine	DZP	11.7	16.4	+40	10.1	15.4	+52
	cc-pVTZ	11.3	16.7	+48	9.8	15.7	+60
L-Glutamine	DZP	12.7	17.7	+39	11.2	16.7	+49
	cc-pVTZ	12.4	17.8	+44	10.8	17.0	+57
D,L-Serine	DZP	14.0	19.2	+37	12.2	18.0	+48
	cc-pVTZ	13.5	19.2	+42	11.7	18.3	+56
L-Threonine	DZP	11.2	14.4	+29	10.0	13.6	+36
	cc-pVTZ	10.9	14.5	+33	9.6	13.6	+42
D,L-Aspartic Acid	DZP	11.6	14.9	+28	10.4	14.3	+38
	cc-pVTZ	11.2	14.9	+33	10.0	14.4	+44
D,L-Histidine	DZP	16.1	22.2	+38	14.1	20.8	+48
	cc-pVTZ	15.6	22.2	+42	13.7	21.3	+55

Table 5 Dipole moments D in Debye from invariom refinement (D_{inv}) and from a refinement of multipole parameters (D_{exp}) using the same multipole model and geometry

Structure	D_{inv}	D_{exp}	$\Delta\mu$ [%]	Theory	Multipole projection	$\Delta\mu$ [%]
L-Alanine	12.1	12.5	+3	11.4	(9.9)	+9
L-Cysteine	11.2	11.2	0.0	10.5	(9.4)	+6
L-Glutamine	13.1	13.4	+2	11.5	(10.8)	+14
D,L-Serine	13.5	12.9	−4	12.5	(11.1)	+3
L-Threonine	11.9	12.0	+1	10.0	(9.2)	+17
D,L-Aspartic Acid	13.1	11.4	−13	10.6	(8.8)	+7
D,L-Histidine	15.7	17.9	+14	14.5	(12.3)	+19

Results of a DFT single-point calculation ("Theory") with the method/basis-set B3LYP/D95 ++(3df,3pd) are given in the right column for comparison. Results from a multipole projection of the DFT density are found to be systematically lower than the single-point results. Hence both single-point (and even more so multipole projection) gives a more pronounced enhancement than the invariom-database [27] fragments

(gas-phase) reference value from theory. A convenient choice for obtaining reference dipole moments for results from experimental multipole refinements is invariom modelling, since it allows a dipole-moment estimate even for large molecules at negligible computational cost. By calculating the difference between experimental dipole moment and the invariom result (1) an enhancement is obtained. To allow a fair comparison between dipole moments from experimental multipole refinement and invariom model we use the same multipole model [7] (i.e. the same local atomic site symmetry and chemical constraints) in both cases. This will be detailed below in the following section.

Invariom modelling is an attempt to apply the benefits of a scattering model that is superior to the independent atom model (IAM) to general small-molecule [33, 59] and ultra-high-resolution macromolecular crystallography [27, 60, 61]. Similar scattering-factor databases are available [38, 39]. In contrast to the experimental multipole refinement, in invariom refinement theoretically predicted multipole populations are kept fixed, so that the number of refinable parameters does not increase with respect to the IAM. Like in the IAM, only positional and displacement parameters are adjusted to the experimental Bragg data.

To put the following results into perspective we need to be aware that both invariom modelling and experimental multipole refinement only permit to obtain the molecular dipole moment within the accuracy the multipole model is capable to provide, as discussed in Sect. 3.

5.1 *Molecular Dipole Moments and Their Enhancement in the Solid State from Experimental Multipole Refinement and Invariom Refinement*

An invariom refinement was performed for the seven datasets considered (Table 1). The input files for invariom refinement were generated by the program INVARIOMTOOL

[40], which also generated input for our experimental multipole refinements. Therefore, the same multipole parameters were adjusted to the experimental data that were used as fixed scattering factors in invariom refinement. Chemical constraints, which are used in the program XDLSM [62] to reduce the number of least-squares parameters in case an identical chemical environment is assumed, were assigned in those cases, where the same invariom scattering factor name was found. Local-atomic site symmetry was chosen in analogy to the model compounds used to generate the database parameters. This way we assured that invariom and experimental multipole refinement were based on the same multipole model. In the multipole refinement, hydrogen atoms were treated as a hybrid scattering factor, where the radial screening parameters κ and the higher multipoles with $l_{max} \geq 1$ were kept at the database values to increase the reliability of the dipole moments obtained (see comments in Sect. 2.1). The invariom geometry was kept. X–H bond distances were set to values obtained in geometry optimizations of model compounds as used in the invariom database [27]. In Table 5 we list the magnitudes of the dipole moments from both invariom and free multipole refinements.

For comparison, molecular dipole moments from a single-point calculation of the experimental geometry are also given. The DFT basis was D95++(3df,3pd) and the functional B3LYP. In analogy to Sect. 3 we include values for the multipole projection of the single-point calculations, which are found to be systematically lower than the values from the single-point calculation. Again, limitations of the Hansen/Coppens multipole model in accurately reproducing dipole moments become apparent.

On the positive side we can see immediately that the extreme spread of values that was observed in a large number of studies [22] is absent. Experimental values are quite close to the theoretical results and reliable estimates from measured intensities are possible following our recommendations on H-atom treatment. However, the accuracy of the multipole model does not allow to clarify whether the enhancement itself is "fact or artefact." This statement is supported by choosing the theoretical single-point dipole moments as reference for assessing a possible enhancement. Since these are systematically smaller than the invariom result, which appears to always yield higher dipole moments than the single-point result, the estimate of the enhancement is also systematically higher (Table 5, right column). These results would be even more pronounced were multipole-projection values (given in brackets in Table 5) of the single-point result taken, which are again systematically lower than the invariom result. Causes for the invariom result giving a higher dipole moment probably lie in the underlying approximation of summing a molecular density from fragments. In conclusion, a more flexible model is needed to answer the question of a possible enhancement. Relying on the answer from theoretical computations (see Sect 4.2) is insufficient, since theoretical calculations predict a pronounced dipole-moment enhancement in all cases in disagreement with experimental findings. We therefore look at results from X-ray constrained wavefunctions in the next section.

Table 6 Dipole moments D in Debye from Hirshfeld-atom refinement (D_{HAR}) and from X-ray constrained wavefunctions (D_{XCW}) from both Hartree–Fock and Density Functional Theory using the in-cluster HAR geometry and the DZP basis set

Structure	HF			DFT		
	D_{HAR}[3]	D_{XCW}	$\Delta\mu$/[%]	D_{HAR}[3]	D_{XCW}	$\Delta\mu$/[%]
L-Alanine	12.6	13.4	+6	11.0	11.9	+8
L-Cysteine	11.7	12.3	+5	10.1	10.7	+6
L-Glutamine	12.7	14.2	+12	11.2	12.8	+14
D,L-Serine	14.0	15.1	+8	12.2	13.1	+7
L-Threonine	11.2	12.9	+15	10.0	12.1	+21
D,L-Aspartic Acid	11.6	12.8	+10	10.4	11.4	+9
D,L-Histidine	16.1	17.1	+6	14.1	15.1	+7

5.2 *Molecular Dipole Moments and Their Enhancement from Hirshfeld-Atom Refinement and Wavefunction Fitting*

Wavefunction fitting [8, 15–18] can be expected to yield better accuracy for properties derived from experimental Bragg data than those derived from the multipole model, since a basis-set description of chosen sophistication can be used to model the electron density. We have chosen the DZP basis [63] already used in Sect. 4.1 (see Table 4). Wavefunction fitting requires a weighting of the experimental data with a multiplier [8] to extract the information content of the individual experimental observations and their standard uncertainties. Hence, the fitting procedure is more demanding than a single-point cluster calculation and needs several repetitions, gradually increasing the multiplier. Geometries obtained from Hirshfeld-atom refinement with cluster charges and dipoles were used and kept fixed. Geometries were assured to be consistent with the basis set this way, which would not have been achieved had invarioms geometries been used. Also, effects on the geometry due to small changes in the dipole moment are avoided.[3] In Table 6 dipole moments obtained are given together with the isolated-molecule result already reported in Table 4. Since the same geometry is used, an enhancement or decrease is reported. A direct comparison to dipole-moment enhancements derived using the Hansen/Coppens multipole model (Table 5) is possible. Analogous to the multipole-model result a strong increase of the in-crystal dipole moment is not observed as it was predicted from theory. Trends from wavefunction fitting hence confirm the results obtained from the multipole model.

[3] A change in dipole moment due to small adjustments of the geometry between Hirshfeld-atom and invariom refinement can be studied by comparing the dipole to the value given in Table 4, where the invariom geometry was used as input. It is found to be insignificant, with the largest difference being 0.1 Debye for Alanine.

6 Discussion: Agreement Between Experimental and Theoretical Results

We would like to obtain an experimental estimate of the dipole moment using as little prior information as possible, since the approximations used in theoretical approaches benefit from independent validation. Unfortunately, experimental data are necessarily limited in resolution. Therefore, a least-squares approach relying on experimental data does not allow an infinite number of parameters to be refined. This restriction leads to an inflexible model and consequently comparably inaccurate dipole moments: when attempting to reproduce the theoretical results limitations of the experimental multipole model approach become apparent.

Such restrictions do not apply to wavefunction fitting, since it combines theory and experiment, using the experimental data as additional information weighted by a multiplier. The quantum-chemical density model is required to fit the experimental data, while simultaneously minimizing the energy of the – now experimental – wavefunction. This allows obtaining a more accurate result at the expense of not providing an entirely experimental result in a strict sense. However, the multipole model also uses a frozen core and fixed radial functions from atomic calculations as input, so that the concept of a purely experimental result from X-ray diffraction seems questionable in general, although this point of view might be considered exaggerated. In spite of such technical details the following results emerge:

1. Accurate in-crystal dipole moments can indeed be obtained from X-ray diffraction.
2. The accuracy of the multipole model is limited, but it can nevertheless provide an estimate of the in-crystal result from experiment after careful modelling.
3. Despite its shortcomings, the multipole-model estimate for the seven experimental data sets studied here is satisfactory. It required taking into account invariom database κ-parameters and optimized X–H distances from model compounds.
4. An estimate of the molecular dipole moment for the crystal geometry can also be obtained entirely from scattering-factor databases without the need for expensive calculations. The invariom result anticipates some of the in-crystal enhancement when compared to single-point calculations for zwitterions.
5. For heavier nuclei (here: S, Cl) the multipole model fails to reproduce dipole moments accurately. Inclusion of κ'-parameter, which can often not be refined in a reliable manner from experimental X-ray diffraction data, is helpful but no remedy for the inaccuracy. Databases can provide κ/κ' values for different chemical environments. Fixing the scale factor in the multipole projection to unity can considerably alter the result, e.g. for sulphur containing compounds.
6. Concerning the enhancement of the dipole moment from experiment in the bulk, and for accurate determinations of dipole moments in general, studies should be preferably based upon a basis-set density representation like it is

used in wavefunction fitting. As a consequence, not only the accuracy but also the computational effort for providing an answer in each particular case is increasing.

7. Dipole moments from cluster calculations consistently predict a substantial in-crystal enhancement. Experimental results, both using the multipole model and wavefunction fitting, suggest a less pronounced enhancement for the amino acids.

7 Conclusion

Seven measurements of high-resolution Bragg data on amino acids published earlier were re-evaluated for a determination of their dipole moments with the Hansen/Coppens multipole model and by a basis-set representation as used in Hirshfeld-atom refinement/wavefunction fitting.

Initially, the general ability of the multiple model to reproduce dipole moments of isolated-molecular calculations was studied by a projection of twenty-two small-molecule electron densities with simulated structure factors. Theoretical dipole moments are usually reproduced within ≈20% of the theoretical result, but can deviate by more than 70% when heavier elements are involved. For the zwitterionic amino acids a systematic underestimation of the dipole moment is seen in the multipole projection. Choices in the treatment of the radial screening parameters κ/κ' as well as the hydrogen-atom scattering are relevant for obtaining a reasonable estimate. Invariom modelling applied on the theoretical geometries – which is also based on the multipole model – equally allows reproducing the dipole moment within a similar range. Here, amino-acid dipole moments are overestimated with respect to the gas phase. On the positive side, the computational effort to obtain dipole moments from database density parameters is minimal. Molecular dipole moments could and should therefore be a routine result of accurate structure determinations. Design choices in the invariom database have been chosen to enable reliable estimates as far as possible.

Refinement of multipole parameters with experimental data allows obtaining the dipole moment of a molecule as part of the crystal. Based on refinements of the seven data sets mentioned, we made suggestions how to make experimental determinations more reliable. Hybrid scattering factors for H-atoms from database approaches and inclusion of accurate optimized X–H bond distances increase the reliability of the determination.

Comparing the experimental dipole and the theoretically predicted invariom moment (or the single-point values) allows assessing dipole-moment enhancements in the bulk, although model inaccuracies limit the reliability of the results. A similar comparison of isolated-molecular calculations and wavefunction fitting using a basis-set representation yields more accurate and consistent results. A density functional theory treatment (BLYP functional) with the DZP basis was performed for that purpose.

To get an estimate on possible dipole-moment enhancements from theory, molecular calculations embedding a central molecule in a field generated by a surrounding cluster of point charges were carried out. These calculations took into account crystal symmetry and used atomic coordinates from invariom refinement. Whereas a considerable enhancement in the bulk is predicted by these theoretical approaches, experimental multipole-model results seem to agree better with isolated-molecular values and do not predict such a considerable enhancement.

To obtain the best possible theoretical estimate for dipole-moment enhancements in the solid state while still taking into account the experimental diffraction data, Hirshfeld-atom refinement within a cluster of point charges and dipoles using density-functional theory and with Dunnings correlation consistent cc-pVTZ basis [64] was performed. Hirshfeld-atom refinement is currently the most sophisticated density model available to refine structural parameters from experimental diffraction data. Theoretical DFT dipole moments allowed putting the experimental results into perspective. The method predicts a significant in-crystal dipole-moment enhancement. However, the extent of the enhancement is a lot lower in wavefunction fitting (5–15% rather than 28–48% for DFT electron densities). It is conceivable that inclusion of molecular van der Waals interactions as provided in dispersion corrected density functionals [65] might bring theoretical estimates and experiment measurements of dipole moments in the solid state closer together.

Our conclusion is that density models more sophisticated than the Hansen/Coppens multipole model increase the reliability of dipole-moment determinations. The accuracy of invariom-database predictions could probably benefit from more accurate density descriptions as well. Wavefunction fitting can currently provide the most accurate experimental in-crystal dipole moment in the presence of high-quality data, albeit at a comparably high computational cost.

Note added in proof A recent experiment shone light on the discrepancy between Hirshfeld-atom refinement within a cluster of point charges and wavefunction fitting of the molecules.

Current program updates in TONTO now allow to perform wavefunction fitting in the presence of surrounding point charges. These lead to an additional enhancement with respect to fitting the molecule only — in better agreement with the experimental data.

We ascribe this to an additional electron density polarization in the vicinity of the core region, which cannot be retrieved with the frozen core in the standard Hansen/Coppens multipole model.

Acknowledgments This work was funded by the Deutsche Forschungsgemeinschaft DFG. B.D. is grateful for support of an Emmy Noether research fellowship, grant DI 921/3-2. We thank R. Flaig, A. Wagner and P. Luger as well as R. Destro and P. Coppens for diffraction data. Fruitful discussions with J. Bak, P. Dominiak, M. A. Spackman and H.-B. Bürgi are acknowledged.

References

1. Stone AJ (1996) The theory of intermolecular forces. Clarendon, Oxford
2. Stewart RF (1969) J Chem Phys 51:4569–4577
3. Kurki-Suonio K (1968) Acta Cryst A A24:379–390

4. Hirshfeld FL (1971) Acta Cryst B 27:769–781
5. Stewart RF (1976) Acta Cryst A 32:565–574
6. Kurki-Suonio K (1977) Isr J Chem 16:132–136
7. Hansen NK, Coppens P (1978) Acta Cryst A 34:909–921
8. Jayatilaka D (1998) Phy Rev Lett 80:798–801
9. Whitten AE, Spackman MA (2006) Acta Cryst B 62:875–888
10. Madsen AØ (2006) J Appl Cryst 39:757–758
11. Munshi P, Madsen AØ, Spackman MA, Larsen S, Destro R (2008) Acta Cryst A 64:465–475
12. Bruning H, Feil D (1992) Acta Cryst A 48:865–872
13. Jayatilaka D, Dittrich B (2008) Acta Cryst A 64:383–393
14. Howard ST, Hursthouse MB, Lehmann CW, Mallinson PR, Frampton CS (1992) J Chem Phys 97:5616–5630
15. Jayatilaka D, Grimwood DJ (2001) Acta Cryst A 57:76–86
16. Grimwood DJ, Jayatilaka D (2001) Acta Cryst A 57:87–100
17. Bytheway I, Grimwood DJ, Jayatilaka D (2002) Acta Cryst A 58:232–243
18. Bytheway I, Grimwood DJ, Figgis BN, Chandler GS, Jayatilaka D (2002) Acta Cryst A 58:244–251
19. Jayatilaka D, Grimwood DJ (2004) Acta Cryst A 60:111–119
20. Palatinus L, van Smaalen S (2002) Acta Cryst A 58:559–567
21. Spackman MA (1992) Chem Rev 92:1769–1797
22. Spackman MA, Munshi P, Dittrich B (2007) ChemPhysChem 8:2051–2063
23. Martin RM (1974) Phys Rev B Condens Matter 9:1998–1999
24. Resta R (1991) Phys Rev B Condens Matter 44:11035–11041
25. El Haouzi A, Hansen NK, Le Hénaff C, Protas J (1996) Acta Cryst A 52:291–301
26. Spackman MA (2005) Phys Rev Lett 95:085502
27. Dittrich B, Hübschle CB, Luger P, Spackman MA (2006) Acta Cryst D 62:1325–1335
28. Coppens P, Abramov Y, Carducci M, Korjov B, Novozhilova I, Alhambra C, Pressprich MR (1999) J Am Chem Soc 121:2585–2593
29. Destro R, Soave R, Barzaghi M (2008) J Phys Chem B 112:5163–5174
30. Destro R, Marsh RE, Bianchi R (1988) J Phys Chem 92:966–973
31. Moggach SA, Clark SJ, Parsons S (2005) Acta Cryst E 61:o2739–o2742
32. Wagner A, Luger P (2001) J Mol Struct 595:39–46
33. Dittrich B, Hübschle CB, Messerschmidt M, Kalinowski R, Girnt D, Luger P (2005) Acta Cryst A 61:314–320
34. Flaig R, Koritsánszky T, Janczak J, Krane H-G, Morgenroth W, Luger P (1999) Angew Chem Int Ed 38:1397–1400
35. Flaig R, Koritsánszky T, Zobel D, Luger P (1998) J Am Chem Soc 120:2227–2238
36. Stewart RF (1973) Acta Cryst A 29:602–605
37. Abramov YA, Volkov A, Wu G, Coppens P (1999) Chem Phy Lett 311:81–86
38. Zarychta B, Pichon-Pesme V, Guillot B, Lecomte C, Jelsch C (2007) Acta Cryst A 63:108–125
39. Dominiak PM, Volkov A, Li X, Messerschmidt M, Coppens P (2007) J Chem Theory Comput 2:232–247
40. Hübschle CB, Luger P, Dittrich B (2007) J Appl Cryst 40:623–627
41. Reid JS (1976) Acta Cryst A 29:248–251
42. Coppens P (1997) X-ray charge densities and chemical bonding, 1st edn. Oxford University Press, Oxford
43. Dittrich B, Hübschle CB, Holstein J, Fabbiani FPA (2009) J Appl Cryst 42:1110–1121
44. Frisch MJ et al (2004) "Gaussian 03, Revision D.01", Technical Report, Gaussian, Inc., Wallingford CT
45. Jayatilaka D, Grimwood DJ (2003) Computational Science – ICCS 2003, 2660:142–151
46. Koritsánszky T, Volkov A, Coppens P (2002) Acta Cryst A 58:464–472
47. Fischer A, Tiana D, Scherer W, Batke K, Svendsen H, Bindzus N, Iversen BB (2011) J Phys Chem B 115:13061–13071

48. Volkov A, Gatti C, Abramov Y, Coppens P (2000) Acta Cryst A 56:252–258
49. Dunmuir DA (1972) Mol Phys 23:109–115
50. Cummins PG, Dunmuir DA (1976) Acta Cryst A 32:847–853
51. Spackman MA, Munshi P, Jayatilaka D (2007) Chem Phys Lett 443:87–91
52. Bader RFW (1990) Atoms in molecules: a quantum theory, vol 22, 1st edn, The international series of monographs on chemistry. Clarendon, Oxford
53. Gatti C, Silvi B, Colonna F (1995) Chem Phys Lett 247:135–141
54. Whitten AE, Turner P, Klooster W, Piltz RO, Spackman MA (2006) J Phys Chem A 110:8763–8776
55. Fantoni AC, Pozzi CG, Punte G (2009) J Phys Chem A 113:9527–9532
56. Dittrich B, Pfitzenreuter S, Hübschle CB (2012) Acta Cryst A 68:110–116
57. Dapprich S, Komáromi I, Byun KS, Morokuma K, Frisch MJ (1999) J Mol Struct (Theochem) 461:1–21
58. Bak KL, Gauss J, Helgaker T, Jorgensen P, Olsen J (2000) Chem Phys Lett 319:563–568
59. Dittrich B, Strümpel M, Koritsánszky T, Schäfer M, Spackman MA (2006) Acta Cryst A 62:217–223
60. Dittrich B, Bond CS, Kalinowski R, Spackman MA, Jayatilaka D (2010) Cryst Eng Comm 12:2419–2423
61. Johnas SKJ, Dittrich B, Meents A, Messerschmidt M, Weckert EF (2009) Acta Cryst D 65:284–293
62. Koritsánszky T, Richter T, Macchi P, Volkov A, Gatti C, Howard S, Mallinson PR, Farrugia L (2003) XD – a computer program package for multipole refinement and topological analysis of electron densities from diffraction data. Freie Universität Berlin, Berlin
63. Dunning TH (1970) J Chem Phys 53:2823–2833
64. Dunning TH (1989) J Chem Phys 90:1007–1023
65. Grimme S (2006) J Comput Chem 27:1787–1799

Struct Bond (2012) 147: 47–98
DOI: 10.1007/430_2010_36
© Springer-Verlag Berlin Heidelberg 2011
Published online: 4 April 2011

Challenging Problems in Charge Density Determination: Polar Bonds and Influence of the Environment

Bernd Engels, Thomas C. Schmidt, Carlo Gatti, Tanja Schirmeister, and Reinhold F. Fink

Abstract The review focuses on the influence of environments on electron densities (ED) and their Laplacians. This is of interest for many applications which uses EDs measured at hand of crystals of a given ligand to predict its pharmaceutical properties. This comprises for example the questions if the ligand fits into the active center of an enzyme and how strong it binds to this active side. This widely used approximation strongly rely on the assumption that the active side of the enzyme influences the ED of the ligand the same way the crystal environment does. This is not obvious since enzymes represent systems made to catalyze reactions. So one could assume that the active sides influence the EDs of ligands in a special way to prepare them for a given reaction. The review shows that this is indeed the case for E64c. Its inhibition properties result since it reacts with cathepsin B and forms a covalently bonded cathepsin B–E64c complex. It clearly comes out that the reaction only takes place since the ED of the ligand is influenced in a way which is not found in the respective crystals. Nevertheless, the review also shows that the above mentioned approximation holds for AMCHA which serves as a model compound for reversible inhibitors. In the last part the review shows in detail that the source function can be used to study the influence of the environment in more detail. In the

B. Engels (✉), T.C. Schmidt, and R.F. Fink
Institut für Physikalische und Theoretische Chemie, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany
e-mail: bernd@chemie.uni-wuerzburg.de

C. Gatti
Istituto di Scienze e Tecnologie Molecolari del CNR (CNR-ISTM), Via Golgi 19, 20133 Milano, Italy
and
Dipartimento di Chimica Fisica ed Elettrochimica, Università di Milano, Via Golgi 19, 20133 Milano, Italy

T. Schirmeister
Institut für Pharmazie und Lebensmittelchemie, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

first part the review summarizes investigations on the reliability of pure theoretical approaches to ED and its Laplacians.

Keywords Electron density · Ab initio · QM/MM · Environmental effects · Source function · E64c · Cathepsin B · AMCHA · Basis set effects

Contents

1 Introduction 48
2 The Accuracy of Theoretical and Experimental Electron Densities 52
3 Environmental Effects on EDs of Biologically Active Molecules 61
4 On the Applicability of the Source Function in QM/MM Investigations 83
5 Summary and Conclusions 89
References 90

1 Introduction

Dalton's atomic hypothesis together with the work of van't Hoff and Le Bel formed the basics to the modern view of matter within which a chemical bonding network links atoms to molecules, aggregates, liquids, or the solid state. This concept is the most widely applied and fundamental model in chemistry and forms, e.g., the basis of the Lewis notation, which provides a reasonable and generally accepted explanation for the electronic and geometric structures of almost all molecules. However, since the 1920s we know that chemical bonding can only be properly understood as a quantum mechanical phenomenon. Furthermore, the work of Kohn and coworkers [1, 2] proved that in principle the electron density (ED) determines all physical and chemical properties of a given molecule [3–6]. The ED is also the underlying quantity in density functional theory (DFT) [7, 8] and readily available from wave-function-based approaches. Its interpretation in terms of bonding effects is possible for example by the quantum theory of atoms in molecules (QTAIM) [9, 10]. This approach allows to quantify the nature of chemical interactions and provides a solid quantum mechanical methodology which allows to interpret the ED. Nevertheless, due to the inherent ambiguity of any interpretation of quantum mechanical properties in terms of localized atomic quantities, QTAIM does not provide an unique interpretation of chemical bonding. Hence, alternatives such as valence bond-based methods [11], energy partitioning approaches [12], population analyses [13, 14], and many others are also important tools.

The ED represents a physical observable as was first noticed by Debye [15]. However, as nicely highlighted in a recent article of Coppens [16], it took much effort to develop an experimental machinery which allows to measure this property with the required accuracy and efficiency. Due to simultaneous developments in experimental setups and analyzing software tools [10], EDs of entire compound classes [17–20] or proteins are today available [19, 21–29], and even more, the routine application of ED measurements is in sight. However, from the literature it

becomes evident that presently the ED, being a formal observable, is hardly measured but reconstructed namely by fitting sophisticated theoretical models to the experimental data, exploiting previous chemical–empirical wisdom [6, 30]. To make this clear in the following, we will talk about "experimentally derived densities" rather than "experimental densities." In principle, experimentally determined electron densities could be used to analyze the shortcomings of theoretical approaches, which, because of the complexity of the systems, use certain approximations in particular with respect to electron correlation and relativistic effects.

However, experimental electron densities can only be used to benchmark theoretical ones if they are only insignificantly biased by the procedures that are used to analyze the measured quantities (i.e., angles and intensity distributions of Bragg reflections). Indeed, various recent studies indicate that such a bias is introduced by the Hansen–Coppens multipole formalism [6, 30], which is most frequently used to derive EDs from experimental diffraction data [31–33]. However, more sophisticated approaches are also biased to some extent [34–36]. This seems to happen because the basis sets used to describe the ED distribution are not flexible enough to reflect details of electron densities in polar bonds. This appears to be at least partly the reason for the differences found between experimentally and theoretically determined topological properties of the ED at the bond critical points (BCPs) which occur in particular at polar bonds, where the densities and Laplacian distributions are strongly different between the bond partners [21–25, 27, 37–39]. For nonpolar bonds, experimentally and theoretically derived bond topological values are in excellent agreement [40, 41]. In most cases, experiment and theory provide qualitatively the same results, for example, with respect to the number of BCPs or the number of valence-shell charge concentrations (VSCCs) [42]. However, there are many cases, for instance, in transition metal complexes or for weak interactions, where the real situation is near a catastrophe point between different topologies, different bond paths, and different numbers of BCPs. In such cases, slightly different experimental or theoretical approaches yield qualitatively different QTAIM-results [6, 43, 44]. A bias of experimental electron densities can also be introduced if two or more parameter sets of the multipole model are of similar quality in residual density and statistical quality. Such a nonuniqueness was demonstrated, e.g., by Peres et al. [45]. Using experimentally derived densities for the validation of theoretical densities is also problematic because the experimental uncertainties – typically 0.1 e/Å^3 for EDs and 4–5 e/Å^5 for Laplacians [17–19, 46–49] – are often of the same magnitude as the differences. Frequently, the computed EDs give even more reasonable trends than the measured ones [50].

However, the disagreement between experiment and theory could also point to shortcomings in the theoretical description. The Hartree–Fock approach for instance is well known for overestimating the polar character of chemical bonds. Thus, only if correlation effects are taken into account reliable densities can be calculated [51–53]. The MP2 approach, however, already predicts topological parameters in close agreement with even more sophisticated approaches such as MP4 or QCISD [32, 54]. Similar densities are also obtained with DFT approaches if

hybrid functionals are used whereupon the results depend on the percentage of exact exchange. Results obtained with gradient-corrected functionals show larger deviations. Also, the influence of the flexibility of the AO basis sets used in the computations was tested to some extent [32, 51]. The observed basis set effects were generally too small to explain the deviations between experimental results and theory. However, in the recent work by Jayatilaka and Dittrich, it was found that the differences between experiment and theory decrease if Slater-type basis functions are used in the computations instead of Gaussian type [55]. In this context it is important to mention that, with the exception of very small model systems, the tested basis sets do not exceed triple-ζ quality although a sufficient number of polarization functions is essential for convergence [32, 50, 51].

Besides the comparison of absolute numbers, an investigation of the correlations between the topological parameters and the bonding type can also reveal important information about the quality of the results. Such correlations are expected and are used in many investigations [3, 4, 6, 38, 56]. One example is the position of the BCP, which, as shown by Cremer and Kraka, correlates with the polarity of the bond [57]. According to Bader's QTAIM approach, correlations are expected between the density, its second derivative along the bond path (λ_3), and the bond ellipticity ε, which is obtained from the second derivatives of the density perpendicular to the bond paths ($\varepsilon = \lambda_1/\lambda_2 - 1$). The bond ellipticity ε can provide information about the delocalization of electrons within a molecule [39, 52]. Results which do not reveal the expected correlations are probably biased to some extent. This is especially true if correlations are found for the less sensitive parameters (e.g., density) but are missing for more sensitive ones (e.g., Laplacian values). For the homoatomic C–C bonds of C_{60} fullerene derivatives, Wagner et al. [41] found the expected correlations between the density at the BCP and the bond distances as well as between the Laplacian at the BCP and the bond distances. This shows that high-resolution synchrotron diffraction experiments provide very accurate densities of these nonpolar covalent bonds. However, whether this also holds true for polar bonds is unclear.

Another uncertainty is caused by the environment. High resolution X-ray experiments are nowadays feasible, but crystals of extremely high purity and quality are required for accurate results. The surrounding of a single molecule in such a crystal differs quite significantly from that in gas phase or within a liquid. One would expect that the environment influences the charge density of the given molecule. Thus, the question arises if densities measured in a crystal can be used to discuss chemical reactivities in solvents, in vacuum or within a protein, e.g., an enzyme. The latter surrounding is of particular importance for rational drug design whose success strongly depends on an intimate knowledge of the interactions between target and active compound [58–62]. According to the Hohenberg–Kohn theorem, these interaction forces are also a function of the ED. High resolution X-ray measurements of the EDs of the enzyme–inhibitor complexes would be an ideal tool to study such interactions as a function of the substitution pattern of the ligands [17–19, 63, 64]. However, high-resolution X-ray diffraction experiments are not yet routine for such large complexes. According to the best of our knowledge, such

datasets are only available for a single protein–ligand complex: the human aldose reductase with the inhibitor IDD594 [65–67].

For the reasons described above, EDs of inhibitors in enzyme–inhibitor complexes can in general only be *approximated* by EDs obtained from crystals of the pure inhibitor compound [17–19, 40, 68–72]. Although the stability of crystals of pure compounds and enzyme–inhibitor complexes is governed by the same types of intermolecular interactions (e.g., steric, electrostatic, hydrogen bonding, van der Waals effects), it is by no means obvious that a crystalline environment is a good approximation for that of an enzyme [73]. Support for this assumption comes from the finding that in enzyme–inhibitor complexes as well as in crystals of the pure compound, the interacting species often arrange in a way that bonding interactions are optimized. Evidence, that the spatial orientations in crystals of small organic molecules are indeed representative for arrangements of inhibitors in enzymes, is given in studies which analyze the Cambridge structural data bank (CSD) to elucidate preferred orientations of functional groups with respect to each other [74]. These orientations were indeed shown to be in qualitative agreement with arrangements in protein–inhibitor complexes from the RCSB protein data bank (PDB) [64, 75–77]. Nevertheless, due to the enormous catalytic power of enzymes, one would expect that the enzyme environment will influence the ED of the inhibitor in a unique fashion. Investigations comparing EDs obtained from crystals of the pure compound and from enzyme–inhibitor complexes are not available. Such investigations may demonstrate to which extent EDs of pure inhibitor crystals can be used to quantify the molecular recognition process between an enzyme and an inhibitor. Insights into the expected differences are also of interest for new approaches which use predetermined atomic EDs to build up EDs of macromolecules, e.g., proteins [45, 78–80]. The quality of the composed EDs might be substantially improved by including the expected influence of the environment in the predetermined EDs [81, 82].

In the present review, we will report on recent results about the influence of environments on the ED of inhibitors [83]. In these works, the EDs of agents in gas phase, polar solvents, crystals, and enzyme environments were compared. These are theoretical studies, since pure experimental approaches are yet not feasible and a mixing of experimental and theoretical values is cumbersome and ambiguous. The quality of the theoretical approach required to achieve this goal is discussed in the first part of this review. As example, we use a series of sulfur–nitrogen bonds possessing a varying degree of polarity. The studies used various theoretical approaches to investigate the convergence of the ED and its topological properties with respect to basis set size and theoretical methods. Furthermore, they focused on the prediction of trends since this is sufficient for correct judgments of the environmental influences.

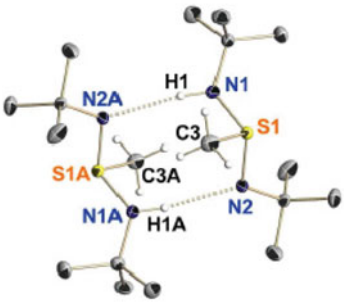
Similar EDs do not necessarily result from similar environmental influences. They may also be caused by a compensation of different effects. This can be investigated using the source function [44, 84] which provides the contribution of different atomic basins on the ED at any given point in space. Hence, in the last part of the review we will report on the applicability of the source function for theoretical approaches that are frequently used to investigate enzyme–inhibitor complexes.

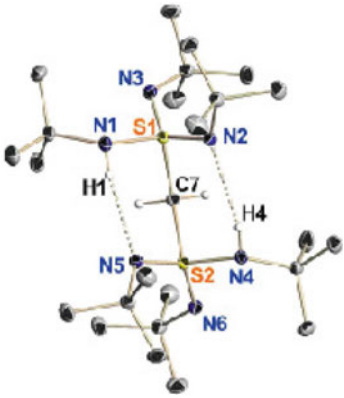
2 The Accuracy of Theoretical and Experimental Electron Densities

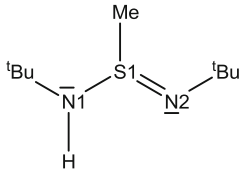
In the next chapter, we will report on theoretical investigations of environmental influences on electron densities. This is only meaningful if the computed electron densities are sufficiently reliable. Thus, in the present chapter we review studies that considered the dependency of absolute values and trends on the theoretical level used.

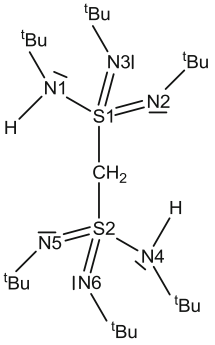
As mentioned before, previous works showed that experimental and theoretical EDs of nonpolar bonds are in excellent agreement. However, less agreement was found for polar bonds. As an example, theoretical and experimental EDs of a set of molecules containing S–N bonds of varying polarity were investigated [17, 50]. The set consisted of methyl(diimido)-sulfinic acid $H(NR)_2SMe$ (**1**), methylene-*bis*(triimido)sulfonic acid $H_2C\{S(NR)_2H(NR)\}_2$ (**2**), sulfurdiimide $S(NR)_2$ (**3**), and sulfurtriimide $S(NR)_3$ (**4**). With R=H, R=Me, and R=*t*Bu, the influence of substituents on the computed densities was investigated. The molecular solid-state structures are given in Fig. 1. In the solid-state structure, the methyl(diimido)sulfinic acid emerges as dimeric unit. To test the influence of the dimerization, a monomeric form of **1** was also computed (**1a**). Similar considerations lead to the investigation of a quasi-monomeric form of **2**, where one methylene-(triimido)sulfonic acid moiety was replaced by a methyl group. This molecule is termed **2a**. Gas phase structures of the model compounds were optimized for different substituents R=H, R=Me, and R=*t*Bu, respectively, using a great variety of theoretical methods. Frequency calculations were performed to ensure that the optimized structures represent stationary points. All calculations were performed with the Gaussian98 package [85].

Tables 1–4 show computed bond topological properties of some typical bonds of the present set of model systems as a function of the method and basis set. The S1=N2 bond of compound **1a** (R=Me, Table 1) was chosen as an example for a formal S=N double bond, while the S1–N1 bond of the same compound was chosen as a formal S–N single bond (Table 2). As a typical example for S–C and N–C bonds to a methyl (or butyl) group, the bond topological properties of the S–C bond of compound **1a** (R=Me, Table 3) were investigated. Table 4 shows the bond topological properties of the formal S1–N1 bond of compound **3** (R=H), which represents another formal S=N double bond, as a function of the basis set. The tables contain the computed bond distance of the given bond, denoted as d, the density, ρ, at the BCP, the value of the Laplacian of the ED at the BCP denoted as $\nabla^2\rho$, and the eigenvalues of the Hessian of the density at the BCP, λ_1, λ_2, and λ_3. In accordance with the usual notation, λ_1 and λ_2 indicate the values associated to the two Hessian eigenvectors that are perpendicular to the bond path direction at BCP, while λ_3 denotes the curvature value along the bond axis. Note that the Laplacian $\nabla^2\rho$ is identical to the sum of these eigenvalues $\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3$. Furthermore, the tables contain the ellipticity $\varepsilon = \lambda_1/\lambda_2 - 1$, and the distances of the BCP to atom A, d(A), and to the other atom, d(B), as well as the ratios d(A)/d(B) and $|\lambda_1|/\lambda_3$. The latter is expected to be smaller than one in ionic bonds [9].

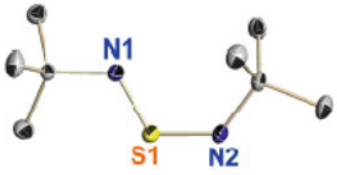


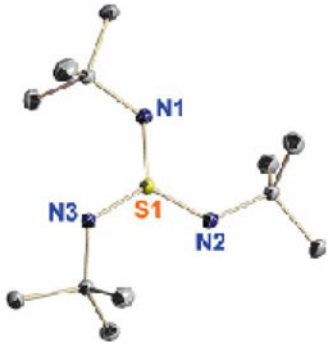


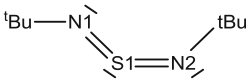


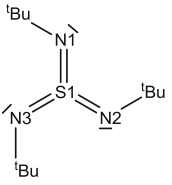


$H(NtBu)_2SMe$ (**1**) | $H_2C\{S(NtBu)_2HNtBu\}_2$(**2**)









$S(NtBu)_2$ (**3**) | $S(NtBu)_3$ (**4**)

Fig. 1 Solid state geometries and Lewis-formulas of **1**–**4**. The anisotropic displacement parameters are depicted at the 50% probability level. In these pictures, CH bonded hydrogen atoms are omitted for clarity

Table 1 Bond topological properties at the BCP of the formal S1=N2 double bond in **1a**, R=Me

–	d	ρ	$\nabla^2\rho$	λ_1	λ_2	λ_3	ε	d(N)	d(S)	$\frac{d(N)}{d(S)}$	$\frac{\lvert\lambda_1\rvert}{\lambda_3}$
B3PW91											
6-31G(d)	1.55	1.76	8.36	−10.01	−7.14	25.52	0.40	0.96	0.60	1.61	0.39
6-31G(d,p)	1.55	1.76	8.33	−10.02	−7.14	25.50	0.40	0.96	0.60	1.61	0.39
6-31G(2d,p)	1.54	1.83	−5.22	−11.25	−8.27	14.30	0.36	0.92	0.62	1.48	0.79
6-31G(3d,p)	1.53	1.86	1.46	−11.65	−8.57	21.68	0.36	0.92	0.61	1.52	0.54
6-31+G(d)	1.55	1.76	8.02	−10.01	−7.15	25.18	0.40	0.96	0.60	1.60	0.40
6-31+G(d,p)	1.55	1.76	7.89	−10.00	−7.15	25.04	0.40	0.96	0.60	1.60	0.40
6-311G(d,p)	1.54	1.80	6.14	−10.31	−7.51	23.96	0.37	0.95	0.60	1.58	0.43
6-311G(2d,p)	1.53	1.86	−6.68	−11.45	−8.44	13.21	0.36	0.91	0.63	1.46	0.87
6-311G(3d,p)	1.53	1.87	0.45	−11.70	−8.61	20.76	0.36	0.92	0.61	1.50	0.56
cc-pVDZ	1.56	1.67	9.25	−8.85	−6.46	24.58	0.37	0.96	0.60	1.60	0.36
cc-pVTZ	1.54	1.86	3.65	−11.02	−7.77	22.43	0.42	0.94	0.60	1.56	0.49
B3LYP											
6-311G(d,p)	1.55	1.79	4.62	−10.31	−7.50	22.44	0.37	0.95	0.60	1.58	0.46
MP2											
6-31G(d)	1.57	1.67	4.40	−9.09	−6.46	19.94	0.41	0.97	0.61	1.61	0.46
6-31G(2d,p)	1.57	1.71	−7.82	−9.98	−7.42	9.58	0.34	0.93	0.64	1.44	1.02
Exp.	1.53	2.24	−9.38	−12.58	−11.73	14.92	0.07	0.74	0.79	0.94	0.84

The geometry was optimized at the indicated level of theory. Distances are given in (Å), densities are given in (e/Å^3), and second derivatives are given in (e/Å^5)

Table 2 Bond topological properties at the BCP of the formal S1–N1 single bond in **1a**, R=Me

–	d	ρ	$\nabla^2\rho$	λ_1	λ_2	λ_3	ε	d(N)	d(S)	$\frac{d(N)}{d(S)}$	$\frac{\lvert\lambda_1\rvert}{\lambda_3}$
B3PW91											
6-31G(d,p)	1.75	1.31	−8.46	−6.94	−6.34	4.81	0.09	0.98	0.77	1.27	1.44
6-31G(2d,p)	1.74	1.30	−6.35	−7.10	−6.45	7.20	0.10	0.94	0.80	1.18	0.99
6-31G(3d,p)	1.73	1.36	−8.49	−7.32	−6.65	5.49	0.10	0.98	0.75	1.31	1.33
6-311G(d,p)	1.75	1.32	−8.58	−7.32	−6.71	5.45	0.09	0.95	0.80	1.19	1.34
6-311G(2d,p)	1.74	1.31	−6.67	−7.37	−6.64	7.34	0.11	0.92	0.82	1.13	1.00
6-311G(3d,p)	1.73	1.36	−8.85	−7.56	−6.86	5.56	0.10	0.95	0.77	1.23	1.36
cc-pVDZ	1.77	1.24	−6.25	−6.06	−5.55	5.36	0.09	0.99	0.78	1.28	1.13
cc-pVTZ	1.73	1.38	−9.95	−7.84	−6.98	4.88	0.12	0.94	0.79	1.20	1.61
B3LYP											
6-311G(d,p)	1.75	1.30	−8.40	−7.32	−6.74	5.67	0.09	0.95	0.80	1.18	1.29
Exp.	1.68	1.76	−7.95	−10.26	−9.66	11.97	0.06	0.83	0.85	0.98	0.86

The geometry was optimized at the indicated level of theory. Distances are given in (Å), densities in (e/Å^3), and second derivatives in (e/Å^5)

The bonds were selected as they possess different polarities. According to a natural bond orbital/natural resonance theory (NBO/NRT) analysis [50] and on the basis of the ratio $|\lambda_1|/\lambda_3$, the S1=N2 double bond of compound **1a** (R=Me, Table 1) possesses a quite polar character. The S1–N1 single bond of the same compound has also a polar character, while the S–C bond of compound **1a** is less polar.

Table 3 Bond topological properties at the BCP of the formal S1–C3 single bond in **1a**, R=Me

–	d	ρ	$\nabla^2\rho$	λ_1	λ_2	λ_3	ε	d(S)	d(C)	$\frac{d(S)}{d(C)}$	$\frac{\lvert\lambda_1\rvert}{\lambda_3}$
B3PW91											
6-31G(d,p)	1.81	1.30	−8.65	−7.42	−7.02	5.79	0.06	0.96	0.85	1.13	1.28
6-31G(2d,p)	1.80	1.27	−6.69	−6.90	−6.50	6.71	0.06	0.95	0.85	1.12	1.03
6-31G(3d,p)	1.80	1.33	−8.98	−7.56	−7.12	5.69	0.06	0.95	0.85	1.12	1.33
6-311G(d,p)	1.81	1.29	−8.09	−7.52	−7.11	6.55	0.06	0.97	0.84	1.15	1.15
6-311G(2d,p)	1.80	1.25	−6.08	−7.01	−6.60	7.53	0.06	0.95	0.85	1.12	0.93
6-311G(3d,p)	1.80	1.32	−8.30	−7.62	−7.19	6.51	0.06	0.95	0.84	1.13	1.17
cc-pVDZ	1.81	1.29	−9.19	−6.99	−6.57	4.38	0.06	0.96	0.85	1.13	1.60
cc-pVTZ	1.80	1.32	−8.32	−7.82	−7.34	6.84	0.06	0.95	0.85	1.12	1.14
B3LYP											
6-311G(d,p)	1.82	1.27	−7.69	−7.44	−7.06	6.81	0.05	0.97	0.85	1.14	1.09
Exp.	1.79	1.54	−8.70	−9.18	−8.72	9.20	0.05	0.99	0.80	1.24	1.05

The geometry was optimized at the indicated level of theory. Distances are given in (Å), densities in (e/Å^3), and second derivatives in (e/Å^5)

Table 4 Bond topological properties at the BCP of the S1–N1 bond in **3** (R=H), calculated with the B3PW91 functional and indicated basis sets

–	d	ρ	$\nabla^2\rho$	λ_1	λ_2	λ_3
B3PW91						
6-31G(d,p)	1.81	1.30	−8.65	−7.42	−7.02	5.79
6-31G(2d,p)	1.80	1.27	−6.69	−6.90	−6.50	6.71
6-31G(3d,p)	1.80	1.33	−8.98	−7.56	−7.12	5.69
6-311G(d,p)	1.81	1.29	−8.09	−7.52	−7.11	6.55
6-311G(2d,p)	1.80	1.25	−6.08	−7.01	−6.60	7.53
6-311G(3d,p)	1.80	1.32	−8.30	−7.62	−7.19	6.51
cc-pVDZ	1.81	1.29	−9.19	−6.99	−6.57	4.38
cc-pVTZ	1.80	1.32	−8.32	−7.82	−7.34	6.84

The geometry was optimized at the indicated level of theory. Distances are given in (Å), densities in (e/Å^3), and second derivatives in (e/Å^5)

The results provided by Tables 1–4 can be summarized as follows. Obviously, for the Pople basis sets the results obtained with the (2d,p) polarization functions deviate considerably from those obtained with other basis sets. But even if we regard this basis as outlier, the Laplacian and the eigenvalues of the Hessian converge slowly with the basis set. In all cases $\nabla^2\rho$ converges worst, for the S=N double bond in **1a** it sometimes even changes the sign. But also the eigenvalues λ_i themselves vary by 20% or more. The values of λ_3 for the S1=N1 double bond of 3 (R=H, Table 4) increase by about 20% going from the cc-pVQZ to the cc-pV5Z basis. For $\nabla^2\rho$ the series cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z yields the values 8.66, 2.62, 0.15, and 0.83 e/Å^5, i.e., even highly demanding calculations with extremely extended basis sets do not provide converged Laplacian values. This is a real nightmare for reliable predictions of absolute numbers. As expected the much smaller Pople basis sets do not show convergence either, even if the (2d,p)

polarization set is taken as an outlier. Additionally, variations of more than 20% were found for topological data obtained with different functionals or the MP2 method. Similar trends were seen for all bonds. A somewhat better convergence was observed for the S–C bond but also in this case the deviations from the experimental data are quite large.

Having this strong dependency in mind, it is not astonishing that the topological data of experimental EDs deviate significantly from the theoretical ones. Indeed, as mentioned before, the Laplacians are very sensitive with respect to the theoretical level. However, the same is also observed for the ED values which usually show much weaker dependencies on basis set or method. In all cases, the experimentally derived densities at the BCPs were found to lie above the theoretical ones. This indicates that in the bonding region the computed density is considerably smaller than the experimental one. It is interesting to note that a similar behavior is found at the position of the nuclei [50]. Nevertheless, the slow convergence indicates problems in the computations [24, 46–49].

The slow convergence of the bond topological data may be due to the complicated electronic structures of the investigated molecules. However, this may not be the case, as other molecular parameters such as the computed bond distances converge as expected. The difference between experiment and theory could stem from the influence of the crystal leading, for example, to slightly different bond distances. To check this possibility, the bond topological data for the experimental geometries were computed. The values are given in Table 5 for the S=N double bond of compound **4**. Also in this case, a very slow convergence with respect to the basis set was found. It is important to note that the deviation from the experiment is even larger if the experimental geometry is used. The difference in the position of the BCPs is quite remarkable. While the experimental BCPs are located almost in the middle of the bond, $d(\mathrm{N})/d(\mathrm{S}) \approx 1$, theory predicts them to be much closer to the sulfur center, $d(\mathrm{N})/d(\mathrm{S}) \approx 1.5$. In a similar case, Gatti and Bianchi improved the agreement between experiment and theory by computing the properties at the same location (e.g., at the experimental BCP) [86]. Similar results were obtained by

Table 5 Bond topological properties at the BCP of the formal S=N double bond of compound **4** with R=*t*Bu (S(N*t*Bu)$_0$) computed at the experimental geometry

–	ρ	$\nabla^2\rho$	λ_1	λ_2	λ_3	ε	d(N)	d(S)	$\frac{d(\mathrm{N})}{d(\mathrm{S})}$
STO-3G	1.49	21.67	−5.83	−3.53	31.04	0.65	0.93	0.59	1.58
SV	1.69	5.75	−7.86	−6.52	20.13	0.21	0.90	0.61	1.47
6-31G(d,p)	1.88	10.61	−11.00	−7.30	28.92	0.51	0.92	0.59	1.57
6-311G(d,p)	1.90	8.09	−11.15	−7.51	26.75	0.49	0.92	0.59	1.56
6-311G(2d,p)	1.94	−7.45	−12.09	−8.20	12.84	0.47	0.89	0.62	1.43
6-311G(3d,p)	1.93	−1.06	−12.09	−8.18	19.21	0.48	0.90	0.61	1.47
6-311++G(d,p)	1.90	8.64	−11.76	−7.54	27.36	0.56	0.92	0.59	1.57
Exp.	2.27	−10.56	−14.40	−11.83	15.69	0.22	0.78	0.74	1.05
6-311++G(d,p)[a]	1.95	−14.28	−11.26	−7.64	4.48	0.47	0.78	0.74	1.05

Distances are given in (Å), densities in (e/Å^3), and second derivatives in (e/Å^5)
[a]Bond topological values at the position of the experimental BCP

Hibbs et al. [19]. In these cases an improved agreement between theoretical and experimentally derived densities and Laplacians were observed. Table 5 shows a similar trend.

The strong deviations between experiment and theory may also result from the replacement of sterically demanding substituents, which are often replaced by smaller groups to lower the computational cost. These substituents are important for the kinetic stability of the substances. It is generally assumed that the electronic structure is not influenced by these groups. For less sensitive properties this assumption is surely justified. For the more sensitive ones such as the Laplacian at the BCP of a polar bond, it has to be tested whether this simplification within theory is still justified. For the present model systems, the bond topological properties were found to be weak functions of the substitution patterns, as shown in Table 6 for the formal S=N double bonds of $S(NR)_3$. Compound **3** was selected as an example since the formal double bonds showed stronger sensitivities with respect to basis set size and method. Similar variations were obtained for all other compounds [50]. Using the 6-311G(d,p) basis set, the Laplacian varies in the series R=H, R=Me, and R=*t*Bu from 3.54 to 3.95 e/Å^5. This variation is smaller than the one resulting from changes in the basis set. Also the ratio d(N)/d(S) remains nearly unaffected. This indicates that the differences in the experimental and theoretical bond topological properties do not result from the different substituents. A replacement of bulky substituents by smaller ones (e.g., R=Me) does not overly influence the results. For the present cases, even R=H seems to be a good approximation for many cases [50].

The strong dependence of the bond topological data on the basis sets may be a result of the fact that the atomic basis sets are centered at the atoms. Hence, the regions around the nuclei are better described than the bonding region where the BCP is located. However, as can be seen from Fig. 2, a different problem seems to exist in the present case. The strong dependence of the λ_3 values on the method of calculation seems to be connected with the position of the BCP. d(N)/d(S) values around 1.5 were computed for the two formal S=N double bonds, S1–N2 of compound **1** (R=Me, Table 5) and S1–N1 of compound **3** (Table 4), i.e., the BCP is located considerably closer to the sulfur than to the nitrogen center.

Table 6 Parameters derived from a linear regression between theoretically and experimentally derived bond topological properties at the BCP, $\rho(r_{BCP})$, and $\nabla^2(r_{BCP})$ vs. bond distance d

–	Calc.	$\rho = a \cdot d + b$			$\nabla^2\rho = a \cdot d + b$		
	–	a	b	R^2	a	b	R^2
Experiment	–	−2.72	6.38	0.91	16.78	−38.47	0.11
B3PW91/6-311G(d,p)	opt[a]	−2.02	4.94	0.98	−97.38	153.03	0.870
B3PW91/6-311G(d,p)	sp[b]	−2.38	5.50	0.939	−28.37	45.35	0.046
PW91/6-311++G(d,p)	sp[b]	−2.24	5.19	0.694	−91.89	144.58	0.633

For the density, eight different S–N and two S–C bonds were selected from compounds **1**−**4**. For the Laplacian values, only the eight different S–N bonds were taken into account

[a]The optimization was performed with the methyl-substituted compounds

[b]Single point (sp) calculations were performed using the experimental (solid state) geometry (R=*t*Bu)

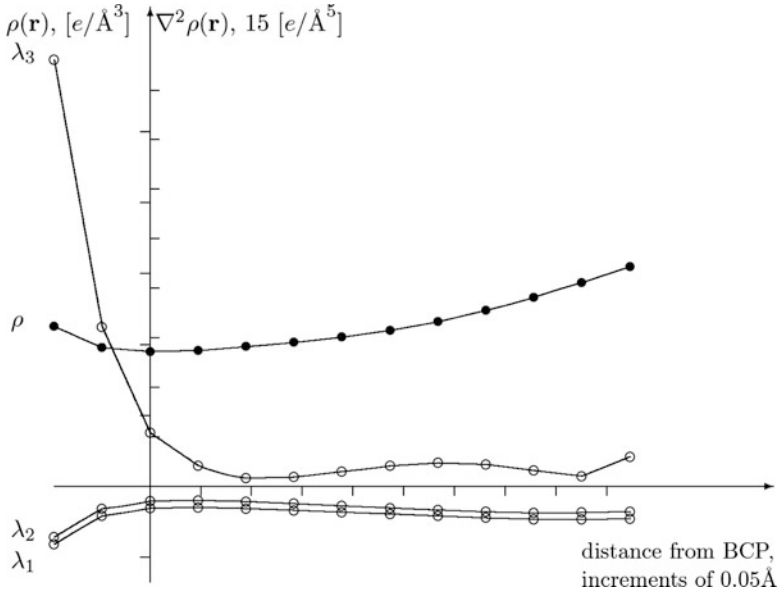


Fig. 2 Eigenvalues λ_i (*empty circles*) and the densities (*filled circles*) along the bond path in **4**, R=*t*Bu, calculated at the B3PW91/6-311++G(d,p) level of theory with fixed experimental geometry. At the BCP (origin of the coordinate system), the density is ρ(BCP) = 1.90 e/Å^3, the Laplacian ($\nabla^2\rho$(BCP)) = 8.26 e/Å^5. The positions of the sulfur and nitrogen centers are at −0.52 Å and 0.92 Å, respectively

Similar values are found for all formal S=N double bonds of the present model systems. For the S1–N1 bond in **1a** (Table 2), ratios around 1.2 were computed, while for the S–C bond in **1a** the ratios are about 1.1. These values are typical for formal single S–N and S–C bonds showing that going from formal S=N double bonds to formal single bonds the BCP moves away from the sulfur center toward the middle of the bonds. The consequence for the density and for its second derivatives with respect to the coordinates can be taken from Fig. 2 in which the density and the λ_i values are plotted along the S1–N1 bond path in compound **4** (R=*t*Bu). The data were calculated at the experimental geometry. For this formal S=N double bond, the BCP is located in a region where λ_1 and λ_2 are almost constant, while the eigenvalue λ_3 varies strongly with the position of the BCP. Thus, the BCP lies in the "rampant edge" of the Laplacian. A tiny displacement of the BCP to the right changes the sign of the Laplacian, while a tiny displacement to the left leads to a strong increase of its positive value. Consequently, already the small variations in the position of the BCP discussed above lead to large changes in the λ_3 value and in the Laplacian. This explains the exceptional results of the (2d,p) set. For the 6-31G(2d,p) and the 6-311G(2d,p), the *d*(N)/*d*(S) values are somewhat smaller than for the basis sets with 1d or 3d polarization functions. For formal single bonds, the variations of the λ_i values along the bond path are similar, but the BCP is located close to the middle of the bonds where λ_3

changes slowly. As a consequence, the value of the Laplacian is much less influenced by details of the calculations.

While some of the differences between theory and experiment will result from uncertainties in the measurements, the data discussed may be interpreted such that computations are yet not good enough to predict topological properties of the density. Hence, we have to ask if theory is appropriate to investigate more detailed influences such as those of the environment. The slow convergence with respect to method and basis sets is found for absolute values of the Laplacian of the density. However, for an investigation of environmental influences, accurate predictions of trends are sufficient. The set of molecules investigated above is ideally suited to explore the capability of theory in this respect. Figure 3 shows the correlation of experimentally and theoretically determined bond distances as well as densities and Laplacians at the BCP. Figure 4 gives the correlation between the densities and the Laplacian at the BCP with the bond distances. Table 6 collects the corresponding linear regression parameters.

As expected, theory and experiment correlate quite nicely for the distances and also for the densities at the BCP (Fig. 3a, b) a moderate correlation is found. The best linear fit for the distances is $y = 1.20x - 0.29$ with a sample correlation coefficient of $R^2 = 0.98$. Experimental and theoretical densities at the BCP can be linearly fitted with the line $y = 0.68x + 0.98$. In this case, the sample correlation coefficient is $R^2 = 0.68$. The S1–N1 single bond of compound **3** represents an outlier in the correlation. No correlation is found between the experimental and the theoretical Laplacians (Fig. 3c).

Figure 4 shows that both the experimental and the theoretical densities at the BCP correlate with the bond distances (Fig. 4a). The linear fit for the experimental data is $y = -3.12x + 7.01$ with a sample correlation coefficient of $R^2 = 0.75$. For the theoretical data, the best linear fit is given through $y = -2.14x + 5.12$ (sample correlation coefficient $R^2 = 0.95$). For the computed Laplacians, a similar but less unambiguous correlation exists ($y = -94.09x + 147.90$; $R^2 = 0.86$). Such a correlation is not found for the experimental counterpart ($y = 16.53x - 38.09$; $R^2 = 0.05$). These data indicate that, at least for the given set of molecules, computations are well suited to investigate correlations. As the present set of molecules is a quite difficult example due to the polar character of the S–N bonds, we expect that this finding can be generalized to other families of compounds. Please note that these calculations were performed with basis sets of moderate size [6-311(d,p)] and the B3LYP functional. This shows that correlations of topological parameters of the ED among themselves and with bond distances do not require very sophisticated theoretical approaches.

The investigation indicates that the correlations within the experimental data are less clear. Taking again into account that the S–N bonds represent quite difficult examples, this cannot be generalized to compounds with less polar bonds. The major reason for the lack of correlations in the experimental data seems to be in this case experimental error bars and uncertainties in the analysis of the experimental data.

Table 5 shows whether experimental geometries can be used for the computations. Although a good correlation between experimental and theoretical values can

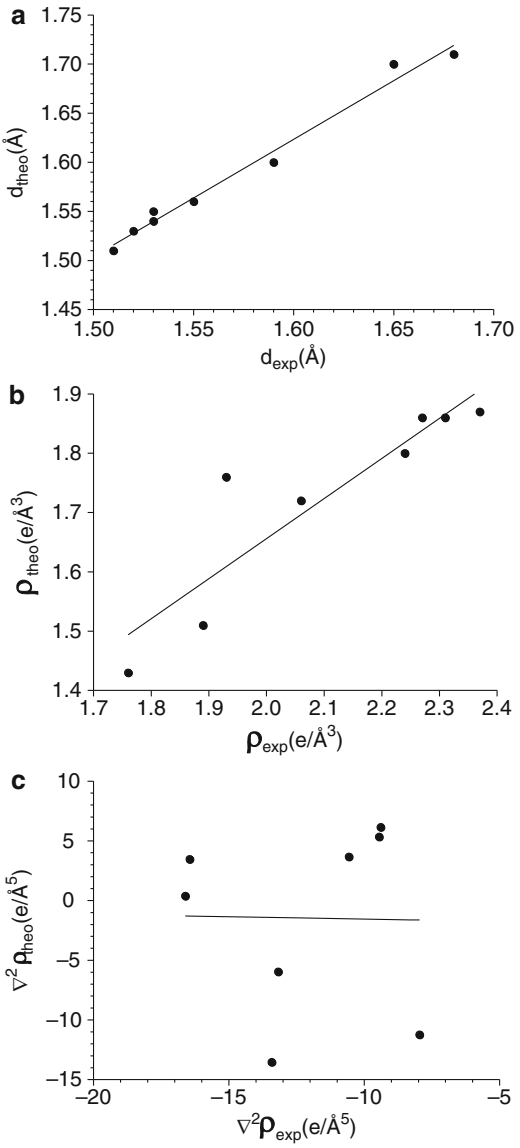


Fig. 3 Correlation between theoretical and experimental data of compounds **1**–**4**. (**a**) Experimental (x-axis) versus theoretical (y-axis) bond distances d of S–N bonds in (Å). Best fit straight line: $y = 1.20x - 0.29$; sample correlation coefficient $R^2 = 0.98$. (**b**) Experimental (x-axis) versus theoretical (y-axis) densities $\rho(r_{BCP})$ of S–N bonds in (e/Å^3). Best fit straight line: $y = 0.68x + 0.98$; sample correlation coefficient $R^2 = 0.68$. The outlier is from S1-N1 in 3. (**c**) Experimental (x-axis) versus theoretical (y-axis) Laplacians $\nabla^2\rho(r_{BCP})$ of the S–N bonds in (e/Å^5). No linear fit possible

be found for small basis sets, in some cases extending the basis sets increases the deviations significantly. However, these deviations may also arise from the properties of the Pople-basis sets (see Tables 1 and 4).

In summary, these investigations showed that a reliable and consistent determination of topological parameters of the ED is still problematic for polar bonds independent of whether they are determined experimentally or with theoretical approaches. The computed data depend strongly on the level of theory, while uncertainties within

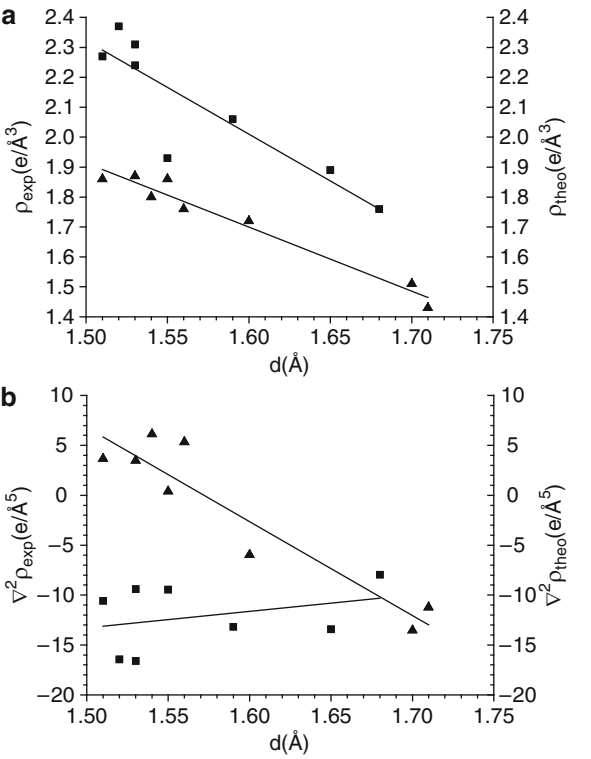


Fig. 4 Comparison of theoretical and experimental correlations. (**a**) Correlation between bond distances d and the corresponding densities at the BCP, $\rho(r_{BCP})$, of S–N bonds. The experimental data are indicated by *squares*, the theoretical counterparts by *triangles*. Best fit straight line for exp.: $y = -3.12x + 7.01$; sample correlation coefficient $R^2 = 0.75$. Best fit straight line for theory: $y = -2.14x + 5.12$; sample correlation coefficient $R^2 = 0.95$. (**b**) Correlation between bond distances d and the corresponding Laplacians of the densities at the BCP $\nabla^2\rho(r_{BCP})$ of S–N bonds. Best fit straight line for exp.: $y = 16.53x - 38.09$; sample correlation coefficient $R^2 = 0.05$. Best fit straight line for theory: $y = -94.09x + 147.90$; sample correlation coefficient $R^2 = 0.86$

the measurements and approximations in the analysis of experimental data limit the reliability of the experimental results. However, trends can be obtained much easier than absolute values. The computed data reveal correlations between the densities at the BCP and the bond distances. Theory also shows a correlation between the corresponding Laplacians at the BCP and the bond distances. The experimental data indicate a correlation between the ED at the BCP and the bond distances, whereas for the Laplacians such a correlation was not found, probably due to experimental uncertainties. Thus, despite the strong dependencies of the absolute values, the following computational approximations are well suited to investigate trends, e.g., the influence of different environments on density-related properties.

3 Environmental Effects on EDs of Biologically Active Molecules

Due to the enormous catalytic power of enzymes, one may expect that the enzyme environment will considerably influence the ED of ligands that are bound to the active site. For rational drug design, the influence of the enzyme environment on

the density of an inhibitor is of great interest. For irreversible inhibitors which deactivate the enzyme by forming a chemical bond to the active site, the influence is expected to be also strong since a chemical reaction takes place. The other type of inhibitors, the so-called noncovalent reversible inhibitors, is attached to the active site by weaker intermolecular interactions which may be too weak to cause a noticeable influence on the ED. In the present text, we review recent theoretical works which investigate whether the ED can be used as an indicator for the inhibition potency of a substance. As discussed above, theoretical data should be sufficiently accurate to reveal correct trends. Using theoretical and experimental data simultaneously is unlikely to be of any advantage as the uncertainties on both approaches are expected to be in the order of or larger than the effects of interest. Additionally, experimental EDs are only available for a single enzyme–inhibitor complex [65–67]. In the following, we will focus on the question whether the electron densities of inhibitors in crystals of the pure compound are comparable to those of the same compound in the active sites of enzymes [83]. The studies required theoretical methods that simulate the influence of different surroundings on EDs on an equal basis, i.e., the inhibitors within the enzyme or the crystal environment. The method of choice to describe such complicated assemblies are combined quantum mechanics/molecular mechanics (QM/MM) approaches [87] since pure quantum chemical approaches are too expensive and MM approaches do not provide an ED. QM/MM methods [88–92] divide the total system (enzyme, solvent, and inhibitor) into the active center and the rest. The active site is described by QM approaches, while the influence of the surrounding protein environment and the solvent is captured at the MM level. The QM and MM regions interact with each other through electrostatic and dispersive terms. In the work reviewed in the following, the electrostatic QM/MM interactions are represented by an electronic embedding scheme [93] incorporating the MM charges into the one-electron QM Hamiltonian and thereby allowing the ED of the QM system to adapt to the field exerted by the environment. Dangling bonds at the QM/MM boundary are capped with hydrogen link atoms [94–98] in the framework of the charge shift method. For the present applications, the inhibitor is described by quantum chemical methods (QM part), while the environment is represented by a force field obtained from molecular mechanical simulations (MM part). These potentials possess atomic resolution, that is, they also contain finer details arising from the molecular nature of the surrounding, for example, due to hydrogen bonds or salt bridges. The geometrical arrangements of the inhibitors in crystals of the pure compound and/or in enzyme–inhibitor complexes were derived from available crystal structures. In addition to crystal and enzyme environments, the influence of polar solvents was also studied using the conductor-like screening model (COSMO) [99, 100]. It should be noted that COSMO is well suited to include the overall polarization by a solvent but has problems to describe stronger interactions, e.g., hydrogen bonding. Completely unpolarized EDs were determined by single molecule computations. This situation corresponds to the inhibitor molecule in vacuum and will be designated as such in the following. The ED of a molecule is a strong function of its geometrical structure. Thus, already different conformers may show strong

variations in the ED. It is possible that such different conformers are found for the structure of inhibitors in crystals of the pure compound and in enzyme environments. Thus, geometries were optimized in different surroundings to differentiate between the direct influence of the environments on the EDs and these more indirect structural effects. The computed data were used for a thorough discussion of the influence of different environment on EDs, electrostatic potentials (ESPs), and multipole moments. The EDs were analyzed with Bader's theory of Atoms in Molecules [9].

The inhibition process of most reversible inhibitors does not involve a chemical reaction, and the noncovalent interactions between inhibitor and enzyme may be too weak for significant variations of the EDs. This is investigated using *trans*-4-(aminomethyl)cyclohexane-1-carboxylic acid (AMCHA, Fig. 5) as an example. It forms a reversible inhibitor–enzyme complex with the recombinant kringle 1 domain of human plasminogen ($K1_{Pg}$) [101]. The preparation of the enzyme–inhibitor complex of AMCHA and $K1_{Pg}$ within the QM/MM framework was obtained as follows: Starting from the 1CEB-crystal structure [101], the influence of the surrounding solvent was incorporated by a water sphere with a radius of 50 Å. Counterions were added to make the system neutral overall. Then a series of consecutive constrained optimizations and further solvation steps were performed until equilibrium structures were reached. The preparation of the system was done by means of fully classical MD simulations using the CHARMM force field and program package [102, 103]. In the next step, the QM/MM approach was applied to optimize the geometrical arrangement of the active site including the inhibitor at a quantum mechanical level. Often a geometry optimization is necessary to obtain reliable information about the ED at the BCP. The resulting structure is depicted in Fig. 6a which includes additionally the numbering of the centers in the AMCHA molecule. The molecular structure of the environment is indicated by sticks, while AMCHA is shown in ball and stick representation. Some important geometrical parameters are collected in Table 7. Finally, the ED of the inhibitor in this reversible enzyme–inhibitor complex was computed. The QM/MM calculations were performed with the ChemShell program [105] using the DL_POLY code [106] for the MM and the TURBOMOLE program suite [107] for the QM part. The EDs were obtained from single-point computations at the DFT/B3 [108]-LYP [109, 110]/TZVP [111] level using the geometries described above. This DFT level is well known to describe many properties with an excellent cost–benefit ratio [112–116], but in other cases fail to provide the right answers [117–120].

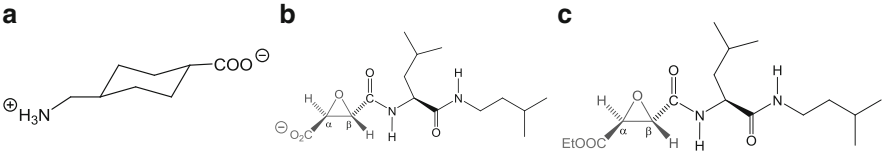


Fig. 5 (**a**) AMCHA, (**b**) Loxistatinic acid (E64c), (**c**) Loxistatin (E64d)

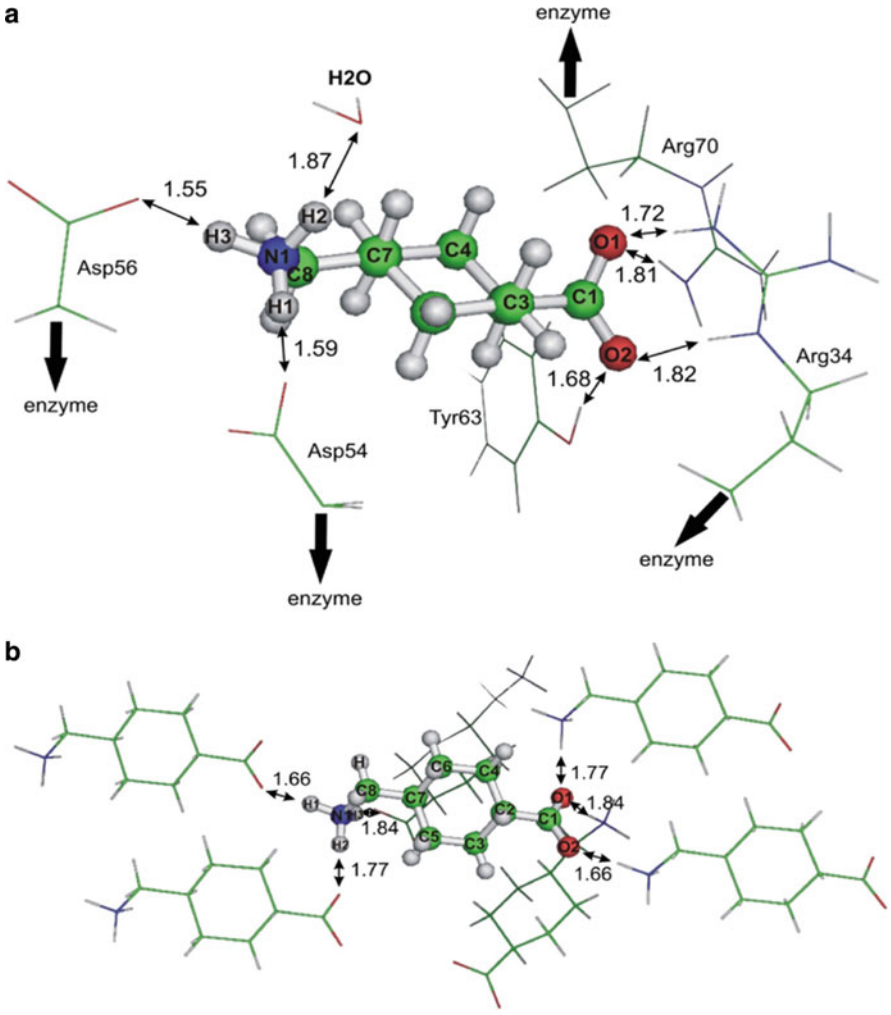


Fig. 6 Hydrogen bonds are indicated by *arrows*. Their lengths are given in Å. (**a**) Orientation of *trans*-4-(aminomethyl)cyclohexane-1-carboxylic acid (AMCHA) in the reversible inhibitor–enzyme complex with the $K1_{Pg}$ (1CEB-crystal structure) [101]. (**b**) Orientation of AMCHA in the crystal of the pure compound [104]

As discussed above, B3-LYP seems to be sufficiently accurate to determine trends in EDs.

Methods that exploit the periodic symmetry of the crystal lattice are well suited to study the influence of crystal environments. However, as the topological properties of the ED and especially its Laplacian values are a strong function of the chosen theoretical approximations, it is not appropriate to compare data obtained from the QM/MM approach discussed above with data computed with periodic boundary conditions. Thus, a QM/MM approach was

Table 7 Selected geometrical parameters of AMCHA in different environments

Parameter	Gas phase	Solvent	Crystal	Protein
C–O1 (Å)	1.260	1.274	1.287	1.287
C–O2 (Å)	1.256	1.277	1.264	1.280
O1–C1–C2 (°)	114.5	118.2	118.6	119.6
O2–C1–C2 (°)	113.4	116.5	116.0	118.9
O1–C1–C2–C3 (°)	95.7	91.1	95.3	91.1
N–H1 (Å)	1.036	1.028	1.068	1.062
N–H2 (Å)	1.036	1.028	1.059	1.040
N–H3 (Å)	1.042	1.028	1.056	1.063
N1–C8–C7 (°)	110.4	112.3	113.7	114.6
N1–C8–C7–C6 (°)	175.0	171.5	168.9	170.7
Average C–C–bond length in hexane ring (Å)	1.546	1.548	1.549	1.549

also used to study the ED in a crystal environment. The X-ray crystal structure of AMCHA [121] was used. While the QM region was chosen as a single AMCHA molecule, the influence of the crystal environment was taken into account by about five shells of surrounding molecules which constituted the MM region. Geometries and EDs were computed and analyzed as for the enzyme–inhibitor complex. The geometrical arrangement is shown in Fig. 6b, which also contains the numbering of the centers of the AMCHA molecule. The environmental structure is again indicated by sticks, while the central AMCHA molecule is shown in ball and stick representation. For a comparison with the enzyme results, some selected geometrical parameters are collected in Table 7.

The influence of a polar solvent was approximated with the COSMO [99, 100] as implemented in the TURBOMOLE program package [107] ($\varepsilon = 78$). These computations were accomplished at the same level of sophistication as for the QM part of the enzyme–inhibitor complex. A comparison of the resulting geometrical arrangement with the structures in the enzyme and the crystal can be taken from Table 7. The different orientations of the carboxylate and of the ammonium group in both environments are obvious, but all other geometrical parameters vary only slightly.

In Fig. 6, the main interactions between the central AMCHA molecule and the environments are indicated by arrows. The corresponding distances of the salt bridges and hydrogen bonds are given in Å. While the interactions seem to be quite similar at the first glance, a more detailed consideration reveals some differences. In crystals of the pure compound, the molecules interact with each other through salt bridges between the carboxylate and the ammonium groups. Each proton of the ammonium groups is involved in one salt bridge. The O1 center of the carboxylate group forms two bridges to hydrogens of the ammonium groups of two different neighbors. The second oxygen of the carboxylate group (O2) interacts with only one hydrogen center of the ammonium group of a third neighbor.

In the AMCHA-K1$_{Pg}$ complex, the ammonium and the carboxylate group also form salt bridges; however, a more detailed view shows some deviations.

The ammonium group forms two salt bridges to the residue of Asp56 (through the H3 center) and Asp54 (through the H1 center). The third proton (H2), however, only builds up a hydrogen bond to a solvent water molecule. The O1 center of the carboxylate group forms two salt bridges with the positively charged guanidinium groups of Arg70 and of Arg34, i.e., its environment is very similar to the one within crystals of the pure compound. As in the crystal, the O2 center is only involved in one salt bridge to the guanidinium group of Arg34. However, a second contact exists to the phenol group of Tyr63. Due to the differences in the networks in crystals of the pure compound and in the enzyme–inhibitor complex, one would expect stronger differences for O2 and H2. On the basis of the numbers of salt bridges and hydrogen bonds, the data for O1, H1, and H3 should be very similar.

Tables 8–14 show the influence of the environments on the bonding properties for the different bonds in AMCHA. The abbreviation *geom/prop* indicates in which environment the geometry was computed and which environment was simulated for the calculation of the property, respectively (g≡gas phase, s≡solvent, p≡protein, c≡crystal). For example, p/s means that the geometry was taken from the protein environment but for the computation of the properties a polar solvent surrounding was modeled. The abbreviation p/p stands for the results taken from the AMCHA-$K1_{Pg}$ complex, while c/c denotes results obtained for the crystal of pure AMCHA. Figure 7 gives contour plot representations of $\nabla^2\rho(\mathbf{r})$ of AMCHA in the plane spanned by the carboxylate group. The geometries are optimized in the respective environment. Figure 8 shows the corresponding Laplacians where the geometry is fixed to the

Table 8 Influence of the environment on topological properties of the density at the bond critical point (BCP) of the C–O2 bond of AMCHA

Geom/prop[a]	ρ	λ_1	λ_2	λ_3
g/g	2.53	−22.31	−20.65	29.09
s/s	2.50	−22.05	−20.93	29.58
p/p	2.43	−20.95	−20.30	25.30
p/g	2.45	−21.18	−19.21	23.04
p/s	2.43	−20.96	−19.82	24.84
c/c	2.49	−21.97	−20.70	29.71
Max:min[b,c]	4.3	6.5	7.5	29.0
p/p:c/c[c]	−2.5	−4.6	−1.9	−14.8
p/p:p/g[c]	−0.8	−1.1	5.7	9.8
p/p:p/s[c]	0.1	0.0	2.4	1.9
p/p:g/g[c]	−6.6	−10.4	−4.2	−19.1
p/p:s/s[c]	−2.7	−4.5	−1.4	−9.2

ρ is the ED at the BCP, while λ_i denotes the eigenvalues of the Hessian of the ED at the BCP. EDs are given in e/Å^3, and Laplacians in e/Å^5. The numbering of the centers can be taken from Fig. 6

[a]The abbreviation geom/prop indicates in which environment the geometry was computed and which environment was simulated in the calculation of the property, respectively (g ≡ gas phase; s ≡ solvent; p ≡ protein; c ≡ crystal). For example, p/s means that the geometry was taken from the protein environment but for the computation of the properties a polar solvent surrounding was modeled

[b][(maximal value − minimal value)/minimal value]*100

[c]Change with respect to p/p in percent

Table 9 Influence of the environments on the bonding properties of the C–O1 bond of AMCHA

Geom/prop	ρ(BCP)	λ_1(BCP)	λ_2(BCP)	λ_3(BCP)
g/g	2.55	−22.67	−20.80	29.76
s/s	2.45	−21.26	−20.21	26.52
p/p	2.38	−20.30	−19.93	24.08
p/g	2.41	−20.67	−18.91	22.03
p/s	2.39	−20.45	−19.35	22.96
c/c	2.36	−20.07	−19.45	23.79
Max:min	8.0	11.5	10,0	35,0
p/p:c/c[a]	0.9	1.2	2.5	1.2
p/p:p/g[a]	−1.0	1.8	5.4	9.3
p/p:p/s[a]	−0.4	0.7	3.0	4.9
p/p:g/g[a]	−6.6	−10.4	−4.2	−19.1
p/p:s/s[a]	−2.7	−4.5	−1.4	−9.2

See Table 8 for further details
[a]Change with respect to p/p in percent

Table 10 Influence of the environment on the bonding properties of the N–H1 bond of AMCHA

Geom/prop	ρ(BCP)	λ_1(BCP)	λ_2(BCP)	λ_3(BCP)
g/g	2.19	−29.78	−29.69	19.81
s/s	2.22	−31.48	−31.43	19.14
p/p	2.02	−28.33	−28.20	17.57
p/g	2.04	−27.18	−27.13	19.38
p/s	2.04	−27.56	−27.49	18.91
c/c	1.99	−27.63	−27.59	17.46
Max:min	9.9	4.9	5.0	13.5
p/p:c/c[a]	1.5	2.5	2.2	0.6
p/p:p/g[a]	−1.1	4.2	3.9	−9.3
p/p:p/s[a]	−0.9	2.8	2.6	−7.1
p/p:g/g[a]	−7.6	−4.9	−5.0	−11.3
p/p:s/s[a]	−9.0	−10.0	−10.3	−8.2

See Table 8 for further details
[a]Change with respect to p/p in percent

Table 11 Influence of the environment on the bonding properties of the N–H2 bond of AMCHA

Geom/prop	ρ(BCP)	λ_1	λ_2	λ_3
g/g	2.18	−29.91	−29.80	19.70
s/s	2.22	−31.33	−31.26	19.32
p/p	2.16	−29.69	−29.56	19.18
p/g	2.16	−29.68	−29.62	19.61
p/s	2.16	−29.99	−29.91	19.20
c/c	2.05	−28.34	−28.27	18.05
Max:min	8.6	8.5	8.6	8.7
p/p:c/c[a]	5.5	4.8	4.6	6.3
p/p:p/g[a]	0.1	0.0	−0.2	−2.2
p/p:p/s	0.2	−1.0	−1.2	−0.1
p/p:g/g[a]	−1.0	−0.7	−0.8	−2.7
p/p:s/s[a]	−2.8	−5.2	−5.5	−0.7

See Table 8 for further details
[a]Change with respect to p/p in percent

Table 12 Influence of the environment on the bonding properties of the N–H3 bond of AMCHA

Geom/prop	ρ(BCP)	λ_1	λ_2	λ_3
g/g	2.15	−29.67	−29.58	19.44
s/s	2.22	−31.39	−31.32	19.24
p/p	2.02	−28.12	−28.07	17.61
p/g	2.05	−26.88	−26.80	19.61
p/s	2.03	−27.44	−27.39	18.91
c/c	2.06	−28.43	−28.37	18.22
Max:min	3.2	13.3	16.8	11.3
p/p:c/c[a]	−2.2	−1.1	−1.0	−3.3
p/p:p/g[a]	−1.3	4.6	4.7	−10.2
p/p:p/s[a]	−0.8	2.5	2.5	−6.9
p/p:g/g[a]	−6.1	−5.2	−5.1	−9.4
p/p:s/s[a]	−9.2	−10.4	−10.4	−8.4

See Table 8 for further details
[a]Change with respect to p/p in percent

Table 13 Influence of the environment on the bonding properties of the C3–C5 bond of AMCHA

Geom/prop	ρ(BCP)	λ_1	λ_2	λ_3
g/g	1.57	−10.37	−10.35	7.68
s/s	1.60	−10.72	−10.67	7.74
p/p	1.59	−10.65	−10.58	7.72
p/g	1.58	−10.48	−10.44	7.68
p/s	1.59	−10.64	−10.58	7.72
c/c	1.60	−10.73	−10.66	7.73
Max:min	1.9	4.7	2.9	0.8
p/p:c/c[a]	−0.5	−0.8	−0.7	−0.1
p/p:p/g[a]	0.8	1.6	1.3	0.6
p/p:p/s[a]	0.0	0.1	0.0	0.0
p/p:g/g[a]	1.3	2.7	2.2	0.5
p/p:s/s[a]	−0.5	−0.7	−0.8	−0.3

See Table 8 for further details
[a]Change with respect to p/p in percent

Table 14 Influence of the environment on the bonding properties of the C4–C6 bond of AMCHA

Geom/prop	ρ(BCP)	λ_1	λ_2	λ_3
g/g	1.58	−10.54	−10.46	7.68
s/s	1.60	−10.75	−10.68	7.74
p/p	1.60	−10.79	−10.69	7.73
p/g	1.59	−10.63	−10.56	7.68
p/s	1.60	−10.78	−10.71	7.74
c/c	1.59	−10.63	−10.53	7.71
Max:min	1.3	2.4	2.3	0.8
p/p:c/c[a]	1.1	1.5	1.6	0.2
p/p:p/g[a]	0.7	1.5	1.2	0.5
p/p:p/s[a]	0.0	0.0	−0.2	−0.1
p/p:g/g[a]	1.3	2.4	2.2	0.7
p/p:s/s[a]	0.2	0.3	0.1	−0.2

See Table 8 for further details
[a]Change with respect to p/p in percent

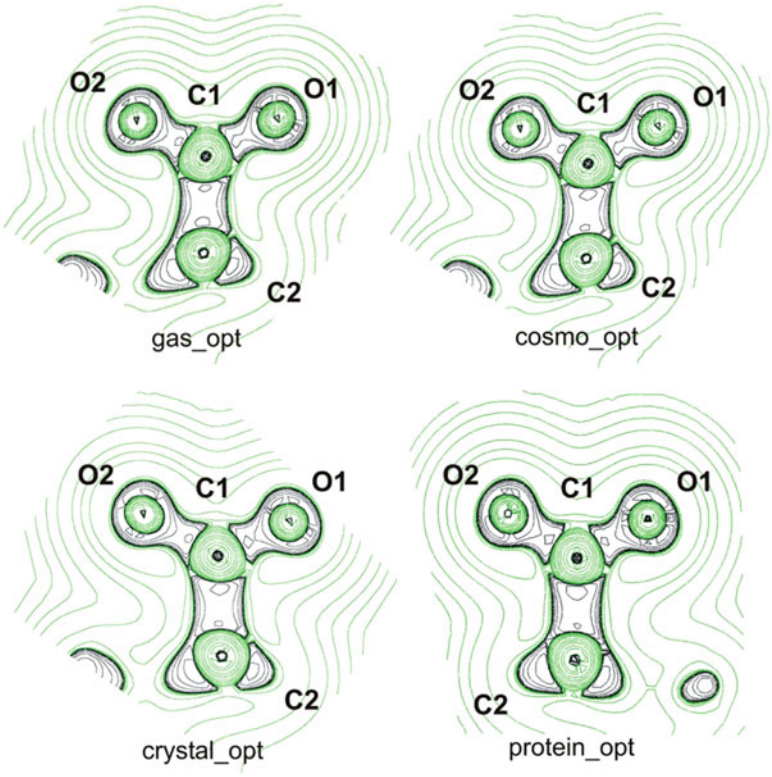


Fig. 7 Contour plot representations of $\nabla^2\rho(r)$ of AMCHA in the plane spanned by the atoms of the carboxylate group (*black lines* negative, *green lines* positive values). All geometries are optimized in the respective environment

situation in the crystal and the enzyme, respectively. This allows differentiating between the variations that arise due to different geometry from those that are due to electrostatic influences. The ESPs are presented in Figs. 9 and 10.

From a simple visual inspection, the differences between the p/p and c/c cases should arise for the C–O2 and the N–H2 bonds. Their environments differ in the kind of the bonding environment or the number of interactions. For N–H2, a salt bridge to a neighbored carboxylate group (in crystal) is replaced by a hydrogen bond to a solvent water molecule (in the enzyme). For C–O2, one salt bridge to an ammonium group (in the crystal) is replaced by a similar salt bridge plus a hydrogen bond to a phenyl group (in the enzyme). The differences are actually reflected in the topological properties of the EDs shown in Tables 8 and 10. The λ_3 values of the C–O2 bond differ for the p/p and c/c calculations by about 15%, while the corresponding difference is only about 1% for the C–O1 bond (Table 9). For the λ_3 value of the N–H2 bond the variation is about 15% (Table 10), while only 2–5% are found for the other N–H bonds (Tables 11 and 12). The corresponding changes at the C–C bonds and C–H bonds within the molecule are smaller than 2% for the ED

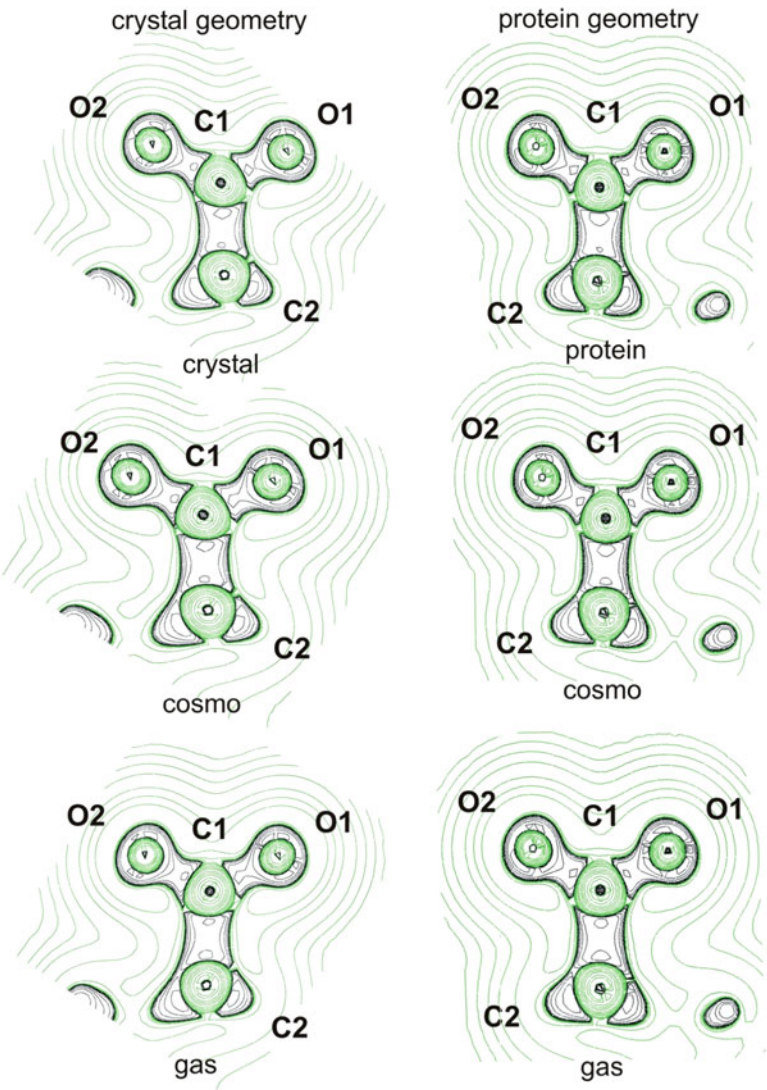


Fig. 8 Contour plot representations of $\nabla^2\rho(r)$ of AMCHA in the plane spanned by the atoms of the carboxylate group (*black lines* negative, *green lines* positive values). The geometries are optimized for the crystal (*left*) and the protein (*right*), respectively. The environment in which the EDs are computed is also indicated

and smaller than 5% for the Laplacians (Tables 13 and 14). Again it is found that the ED is less affected than the Laplacian and the related eigenvalues of the Hessian of the density. This underlines that the ED is a rather insensitive property.

The results from above show that the calculated topological properties of the ED exhibit small but systematic changes due to the environment. Thus, the question

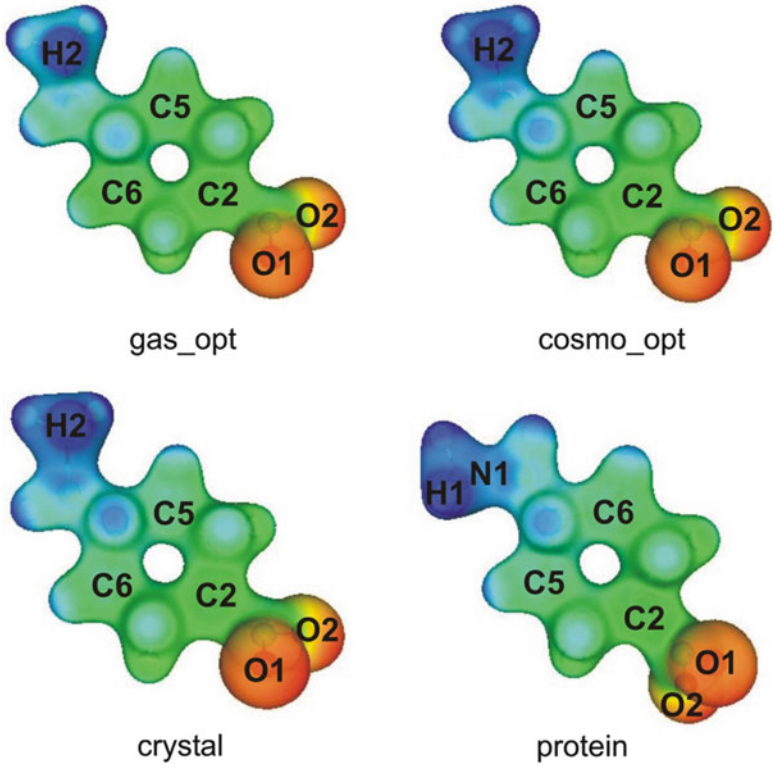


Fig. 9 Electrostatic potentials of AMCHA as computed in the indicated surroundings. All geometries were optimized in the respective environment

arises whether such differences can be measured experimentally. According to Luger, the experimental uncertainties are about 0.1 e/Å^3 for EDs and 4–5 e/Å^5 for Laplacians [24]. In the calculations, the largest absolute differences are found for the λ_3 curvature of the C–O2 bond. The computations predict 25.3 e/Å^5 for the p/p environment, while 29.7 e/Å^5 is computed for c/c. This difference is just as large as the experimental error bars. Taking into account that the determination of EDs in protein crystals is substantially more difficult, it is obvious that this difference is presently much too small to be detected experimentally. Note that all other differences are even smaller. In summary, experimental EDs of the reversible inhibitor–enzyme complex between AMCHA and $K1_{Pg}$ and of the AMCHA-crystal are predicted to be virtually identical. This is also seen if the contour plots (Figs. 7 and 8) and the ESPs (Figs. 9 and 10) of the c/c and the p/p situation are compared.

While the c/c and p/p situations are essentially identical, larger differences are seen if both datasets are compared with pure gas phase results (g/g). Let us concentrate on the C–O1 bond that forms two salt bridges within crystal and protein environments. The λ_3 values obtained for the p/p and the g/g situation differ by about 20% (Table 9). If the ED is computed at the protein geometry for a gas phase

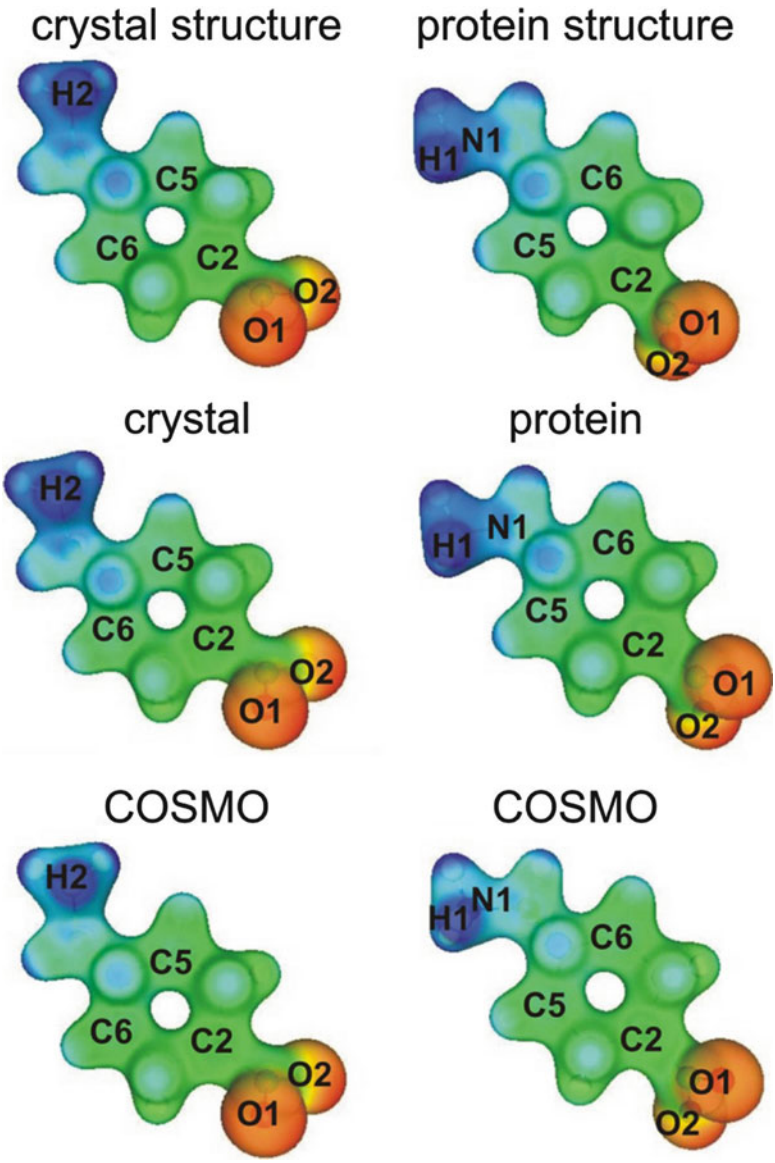


Fig. 10 Influence of the environment on the electrostatic potential of AMCHA. The geometries were optimized for the crystal (*left*) and the protein (*right*), respectively. The environment in which the EDs were computed is also indicated. The last row gives gas phase values at different geometries

environment, the difference reduces to about 9% showing that both geometry and environment influence the density properties in a similar manner. The absolute difference (24.1 vs. 29.8 e/Å^5) is again smaller than the experimental error bars.

An excellent approximation for the protein environment is the p/s situation. It uses the geometry of the enzyme–inhibitor complex but approximates the environment through a continuum approach (COSMO, $\varepsilon = 78$) which simulates a polar solvent. For the C–O2 bond, this approach deviates even less from the p/p results than the data computed for crystals of the pure compound (c/c). In all other cases both approaches are very similar. Their deviations from the protein data are well below 10%.

The computed dipoles (Table 15) reveal similar trends as the bonding parameters. Compared to the gas phase (g/g), we compute dipole enhancements of 46% for the protein environment and 37% for the surrounding inside of crystals of the pure compound. The latter increase resembles values compiled in a recent overview from Spackmann et al. [20] and matches those found by Gatti et al. for urea [122] crystals, which, analogously to AMCHA, contains C=O and NH_2 groups interacting through H-bonds. The environment inside crystals of the pure compound mimics the influence of the protein surrounding very accurately despite the differences in the geometries of AMCHA. As already found for the ED and its Laplacian, the p/s situation represents an excellent approximation to the p/p computation. A comparison between s/s and p/s shows that the strong dipole enhancement in the protein results partly from the geometry of AMCHA in the protein.

Table 16 offers a more detailed picture about the influence of the environments on the charge distribution of AMCHA. The table collects atomic charges computed with the Roby–Davidson partitioning [123, 124]. If the gas phase is compared with the protein environment, a slight increase in the absolute values of the partial charges of the carboxylate and the ammonium group can be found. In this case the partial charge of the carboxylate changes from −0.56 to −1.02, i.e., for about 0.5 units. An increase of about 0.1 is found for the ammonium group. These changes are consistent with the results on the ESPs (Figs. 9 and 10). The increase due to the protein environment is strongest, but the surrounding in crystals of the pure compound and the polar solvent environment yield comparable enhancements, i.e., both represent very good approximations of the protein surrounding.

As also found for enzyme–substrate reactions, the inhibition mechanisms of irreversible inhibitors include covalent bond formation and rupture. Thus, it seems likely that enzyme environments influence EDs of irreversible inhibitors in a similar way as they do for substrates. However, the enzyme environments have

Table 15 Influence of the environment on the dipole moment of AMCHA

Geom/prop	Dipole moment/a.u.
g/g	10.6
s/s	14.1
p/p	15.5
p/g	11.1
p/s	14.5
c/c	14.5
c/g	11.1
c/s	14.0

All values are given in (a.u.). An explanation of the abbreviations can be taken from Table 8

Table 16 Influence of the environment on atomic charges (Roby–Davidson scheme) [123, 124] of AMCHA

Center or group	g/g	s/s	p/p	p/g	p/s	c/c
O1	−0.43	−0.44	−0.52	−0.32	−0.46	−0.53
O2	−0.56	−0.50	−0.50	−0.40	−0.49	−0.46
C1	0.43	−0.02	−0.01	−0.02	−0.03	0.03
Carboxylate	−0.56	−0.96	−1.02	−0.74	−0.98	−0.95
N1	0.13	0.18	0.06	0.11	0.15	0.04
H1	0.13	0.16	0.22	0.15	0.16	0.22
H2	0.16	0.15	0.15	0.14	0.16	0.20
H3	0.14	0.15	0.22	0.14	0.17	0.19
Ammonium	0.56	0.64	0.66	0.54	0.64	0.66
Cyclohexane ring	−0.11	0.23	0.23	0.14	0.20	0.22

All values are given in (a.u.). The numbering of the centers can be taken from Fig. 1; further abbreviations from Table 8

Scheme 1 Two-step model for irreversible inhibition of enzymes. *E* enzyme, *EI* reversible, noncovalent enzyme–inhibitor complex, *E–I* irreversible, covalent enzyme–inhibitor complex

been optimized for the reaction with the substrate in the evolutionary process which is not the case for irreversible inhibitors. Their inhibition potency may result from their inherent chemical reactivity and less from the influence of the enzyme. In this case, the influence of the enzyme environment on the EDs could be as small as for reversible inhibitors. As first test examples, we chose E64-derived compounds (Fig. 5). They are irreversible inhibitors of cysteine proteases, which represent promising drug targets for osteoporosis [125], arthritis [104], cancer [126], and Alzheimer's disease [125]. They react in a two-step mechanism (Scheme 1) [127, 128]. In the first step, a reversible noncovalent enzyme–inhibitor complex EI is formed. The stability of this complex arises from reversible interactions between ligand and enzyme. The second step is initiated by the nucleophilic attack of the negatively charged Cys thiolate at C_α of the epoxide moiety. Subsequently, a ring opening reaction happens in which a new S–C_α bond is formed while the C_α–O bond of the epoxide is broken.

The investigations for the E64c compounds started from the crystal structure of the covalently bonded cathepsin B–E64c complex (1ITO) [129], which is the final product of the two-step mechanism (Scheme 1). The preparation of the system was performed as described for AMCHA. The resulting structure is shown in Fig. 11. The molecular structure of the environment is indicated by sticks, while E64c is shown in ball and stick representation. The numbering of the centers of E64c is also given. The main interactions between E64c and the active site of cathepsin B are indicated by arrows.

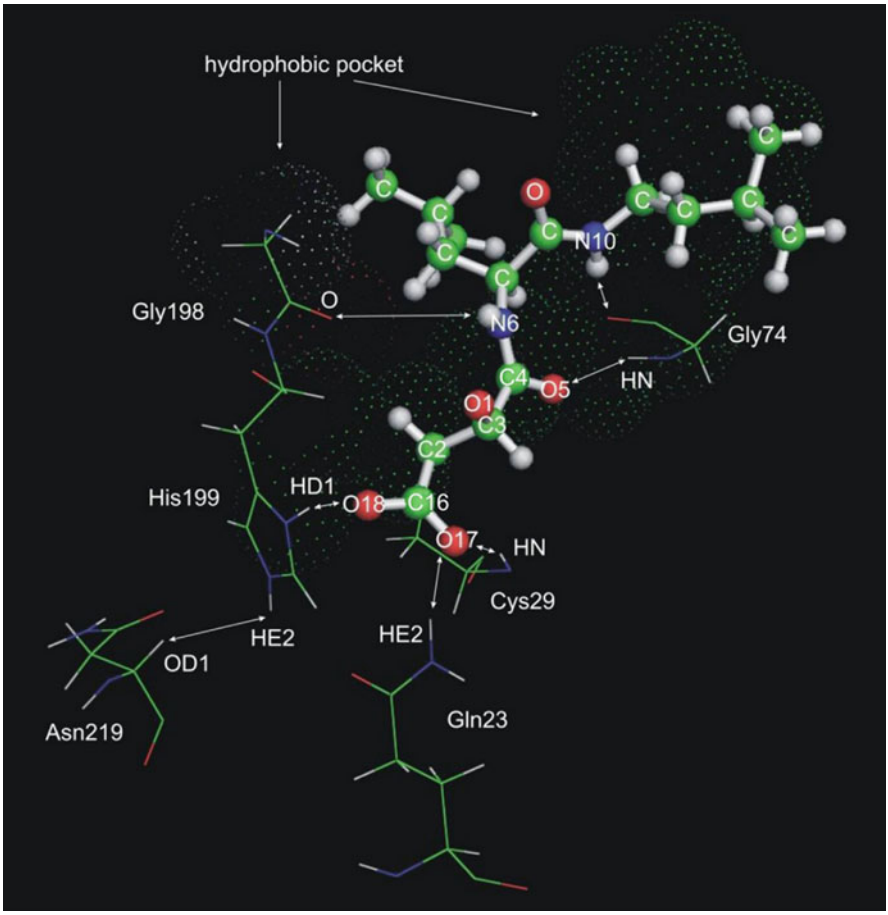


Fig. 11 Prepared structure of the irreversibly inhibited cathepsin B enzyme. The preparation started from the 1ITO X-ray structure [129]

However, not the irreversible but the reversible enzyme–inhibitor complex EI (Scheme 1) must be considered as reactant of the inhibition reaction. It is needed for the investigation since, according to the Hammond postulate, the energetically higher reactant tells us more about the transition state and, thus, about the reactivity, than the product. However, X-ray data from this complex are not available since the epoxide moiety is too reactive. To determine the reversible enzyme–inhibitor complex, we computed the reaction path backwards from product (irreversible complex) to the reactant (reversible complex). QM/MM reaction path calculations were performed using DFT at the RIDFT [130] B [108]-LYP [109, 110]/TZVP [111] level for the QM part and the CHARMM force field for the MM part. In these computations, the QM part comprises the inhibitor and the residues of the Cys29 and of the His199 moieties. The rest of the active site is in the MM-part.

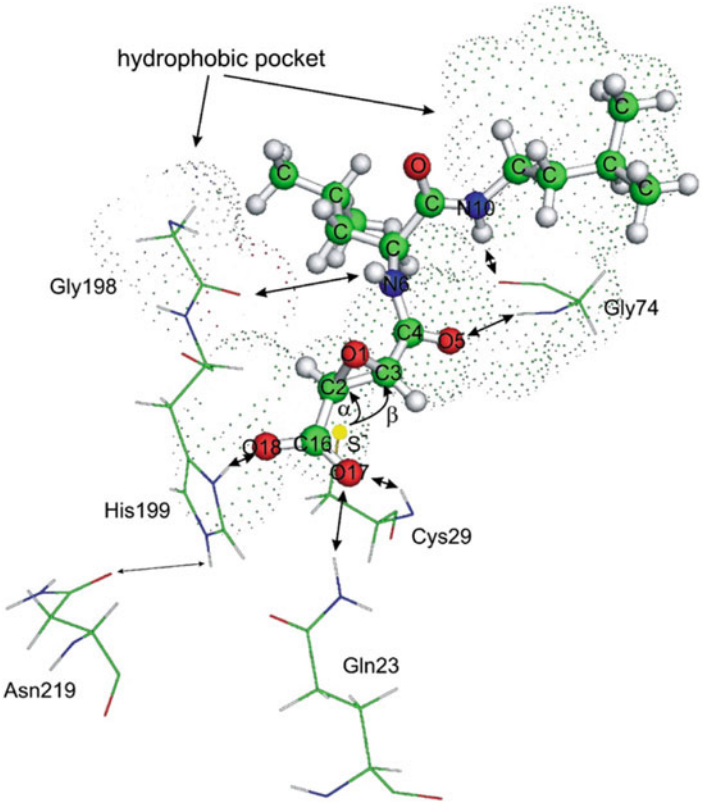


Fig. 12 Structure of the reversible enzyme–inhibitor complex. As indicated, the thiolate may attack at the C_α and C_β (C2 and C3) carbon atoms

The resulting structure of the reversible enzyme–inhibitor complex is given in Fig. 12. In this figure, we use the same representation as in Fig. 11.

The structure of the computed reversible inhibitor–enzyme complex cannot be directly compared with experimental data as X-ray experiments are not feasible. However, the computed data for the inhibition reaction agreed with all relevant experimental results about the regio [131] – and the stereochemistry [132] of the inhibition process, the influence of the pH value of the environment [133], and trends in the potencies of the inhibitors. The agreement for this multitude of properties strongly supports the reliability of the computed structure. Further details on the preparation procedure and the computations of the reaction paths are described elsewhere [134].

In the next step, the QM/MM approach was applied to compute the EDs of the inhibitor in the reversible enzyme–inhibitor complex. The calculations were performed as for AMCHA. For a comparison of the situation within the enzyme with the one in vacuum and solvent, E64c was also computed for these environments. The geometrical arrangements are given elsewhere [83].

Unfortunately, the crystal structure of the pure E64c compound is not available. Hence, we used the known crystal structure of E64d [135] to compare the crystal environment with the enzyme surrounding. In biological systems, E64d is the precursor that is hydrolyzed by esterases to the active compound E64c [127, 128]. Thus, for E64d X-ray scattering data of the enzyme–inhibitor complex are not available.

As indicated in Fig. 12 the ring opening reaction can be induced by an attack of the thiolate group of the Cys29 moiety at the α- or the β-carbon center of the epoxide ring. In a polar solvent, the β-attack is preferred for epoxide carboxylates reacting with phenylthiolate. Within the cathepsin B environment, E64c reacts only via α-attack, i.e., the enzyme environment flips the regioselectivity of the reaction.

As already discussed previously [131], strong interactions between E64c and the active site of cathepsin B result from the carboxylate group (C16–O17–O18) which interacts with the oxy-anion hole (residue of Gln23 and backbone NH-group of Cys29) and the protonated His199 residue. These interactions have a strong effect on the kinetics of the inhibition reaction since they pull the inhibitor into the active site toward the attacking thiolate group of the Cys29 residue. As can be seen from the geometrical data summarized in Table 17, the distance between the reacting species becomes so small that the ring already starts to open in the reactant valley (see the distance C2–O1 and the angle O1–C3–C2). In comparison to gas phase or solvent data in the enzyme–inhibitor complex the C2–O1 bond is elongated and the angle O1–C3–C2 is enlarged indicating a strong distortion of the reactant structure toward the product arrangement. The single interactions connected with the peptidomimetic side chain are considerably smaller but in sum lead to a more pronounced regiospecificity [131] and determine the stereoselectivity [132] of the inhibition process.

According to the AIM theory developed by Bader and co-workers [9], regions with $\nabla^2\rho(r) > 0$ represent domains with charge depletion. Upon chemical combination, the atomic region of VSCC, with $\nabla^2\rho(r) < 0$, may be punctured and locally destroyed, leading to the formation of regions of relative charge depletion within the VSCC, often characterized by positive Laplacian values. The (3,+1) $\nabla^2\rho(r)$ ring critical points associated with the triangles formed by (3,−3) bonded or non-bonded

Table 17 Selected geometrical parameters of E64c in different environments

Parameter (E64c)	Gas phase	Solvent	Protein
C16–O18 (Å)	1.27	1.27	1.28
C16–O17 (Å)	1.26	1.27	1.25
C2–O1 (Å)	1.47	1.47	1.53
C3–O1 (Å)	1.43	1.45	1.45
O18–C16–C2 (°)	111.6	113.9	117.7
O17–C16–C2 (°)	116.9	118.3	119.1
O18–C16–C2–C3 (°)	−128.7	−128.7	166.8
O1–C2–C3 (°)	57.7	58.5	58.1
O1–C3–C2 (°)	60.0	59.8	64.0
C16–C2–C3–C4 (°)	141.9	147.5	156.3

charge concentrations as vertices are typically characterized by less negative or even positive Laplacian values and represent the positions where a nucleophile will attack with highest probability. The $\nabla^2\rho(r)$ value at the (3,+1) critical point with higher and positive Laplacian thus provides a measure for the electrophilicity of the atomic center to which it is related and can be used to discuss trends.

Test computations for a smaller model system **8** consisting of the epoxide and the carboxylate substituent (Fig. 13 left-hand side) reveal that the orientation of the carboxylate group influences the $\nabla^2\rho(r)$ values at the (3,+1) critical points significantly (Fig. 14). The values computed for the geometry taken in the E64c–cathepsin B complex differ from the corresponding values in the solvent since the enzyme –inhibitor interactions lead to additional distortions in the epoxide ring (Table 17).

The inhibitor adopts different geometries in different surroundings. To distinguish between the direct influence of the environment on the ED and more indirect effects

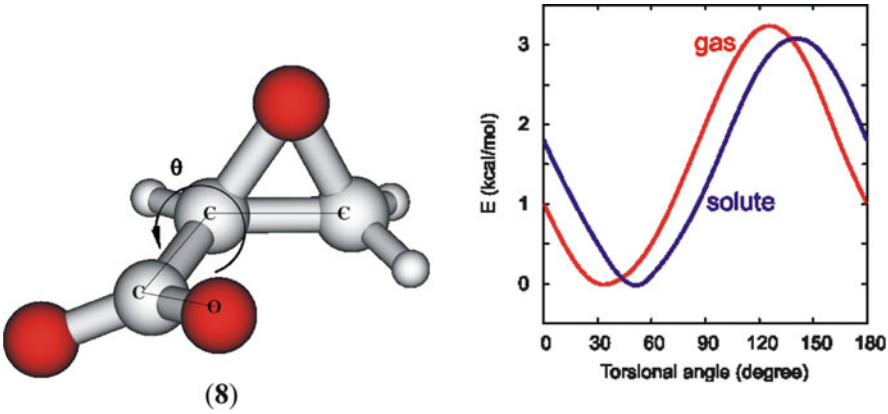


Fig. 13 Potential energy curve for the rotation of the carboxylate group by the torsional angle θ. The computations were performed for the substituted epoxide **8**

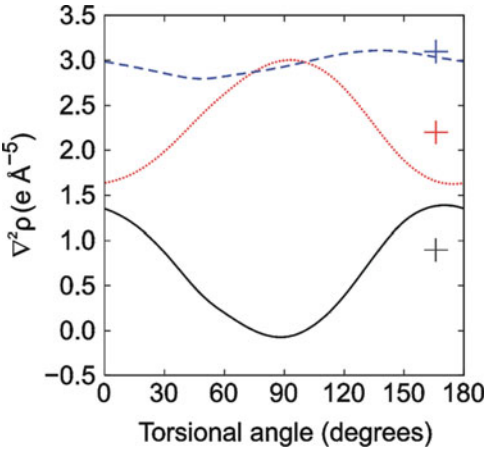


Fig. 14 Variation of the Laplacian value of **8** (Fig. 13) at the (3,+1) critical point at the α (*red*) and β center (*blue*) as a function of the torsional angle θ. The computations are performed for a polar solvent. In the solvent, the equilibrium geometry is found for $\theta = 55°$. The *black line* is the difference between the Laplacians values of the α and β centers. The *crosses* depict the values obtained for the protein structure

which are caused by the different geometries, the E64c molecule was fixed at its geometry in the enzyme–inhibitor complex, and the ED and its derivates were computed for gas phase (p/g), polar solvent (p/s), and two different enzyme environments. The abbreviation p/p(S^{-1}) denotes the real situation in the enzyme, i.e., the thiolate is negatively charged. For p/p(S^0), the charge of the thiolate is artificially set to −0.07 e. This variation allows analyzing how such a strong change in the environment influences the ED of the inhibitor. Figure 15 shows the corresponding contour plot

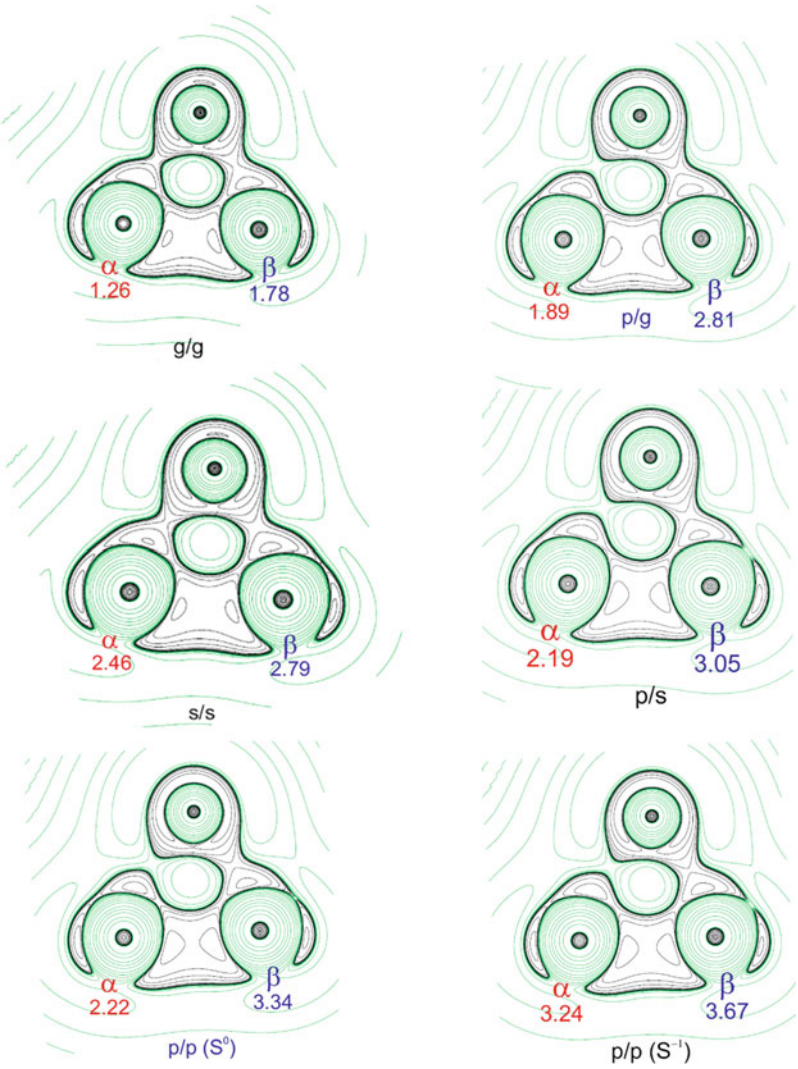


Fig. 15 Contour plot representation of the Laplacian of E64c in the epoxide plane. The values of the (3,+1) critical points are given. Abbreviations for geometry/environment can be taken from Table 16

representations of the Laplacian of E64c in the epoxide plane and gives the values of $\nabla^2\rho(r)$ at the (3,+1) critical points. For comparison, we also included the values obtained if geometry and ED are determined in gas phase (g/g) or polar solvent (s/s), respectively.

While Fig. 15 shows influences of the environment on the values of $\nabla^2\rho(r)$ at the (3,+1) critical points, the general shape of $\nabla^2\rho(r)$ does not change significantly in comparison to data obtained for the protein geometry. Additionally, the sequence of the electrophilicity of the carbon centers is not changed, i.e., C_β always represents the more electrophilic center. The $\nabla^2\rho(r)$ values at the (3,+1) critical points increase significantly if the gas phase environment (p/g) is replaced by a polar solvent (p/s) or if the charge of the thiolate is changed from -0.07 (p/p S^0) to -1 (p/p S^{-1}). The latter effect owes to the repulsion of the negatively charged thiolate and the electron cloud of the epoxide ring. Since the thiolate-C_α distance is smaller, the $\nabla^2\rho(r)$ value at C_α increases stronger than the one at C_β. As found for AMCHA, the polar solvent mimics the protein environment considerably better than the gas phase approach.

The distortion of the epoxide ring induced by the protein environment is clearly reflected in $\nabla^2\rho(r)$. The C_α–O1 bond (≡C2–O1 bond) is considerably weakened. The C_β–O1 bond, on the other hand, is not affected. Combining the present data with previous computations [131, 132] about the kinetics, a clear picture emerges as of how the interaction in the enzyme–inhibitor complex influences the course of the subsequent reaction. Due to the strong bonding interactions between the carboxylate group and the oxyanion hole and the His199 residue, the epoxide ring is pulled so close to the thiolate that the epoxide ring is strongly distorted toward the product structure. Due to the position of the thiolate and the strongly weakened C_α–O1 bond, only the α-attack takes place despite the higher electrophilicity found for C_β. This effect is strengthened by the interactions between the enzyme and the peptidomimetic side chain. They clamp E64c into the active site so that the β-attack is additionally disfavored. In total a reaction barrier of only 1–2 kcal/mol was computed for the α-attack, while for the β-attack a barrier or 15 kcal/mol was predicted. Within the picture of a potential energy surface, the regiospecificity is caused by the fact that the reactants are pushed along the reaction coordinate toward the transition state. Looking at the ED of the reactants, the significantly elongated C_α–O1 bond is one of the main reasons for the regiospecificity. Such special influences can, of course, not be mimicked in pure gas phase (g/g) or solvent computations (s/s). Most probably it is also not possible to deduce them from measurements performed for crystals of the pure compound.

X-ray measurements of the crystal structure of E64c are not available. Hence, to get a rough insight into the influence of the crystal environment on the electronic structure of the epoxide moiety, we used X-ray data of E64d (Fig. 5). In E64d the carboxylate group of E64c is replaced by an ester moiety. The geometrical arrangements of the units in crystals of pure E64d are depicted in Fig. 16. Figure 17 summarizes the corresponding contour plot representations of $\nabla^2\rho(r)$ and the

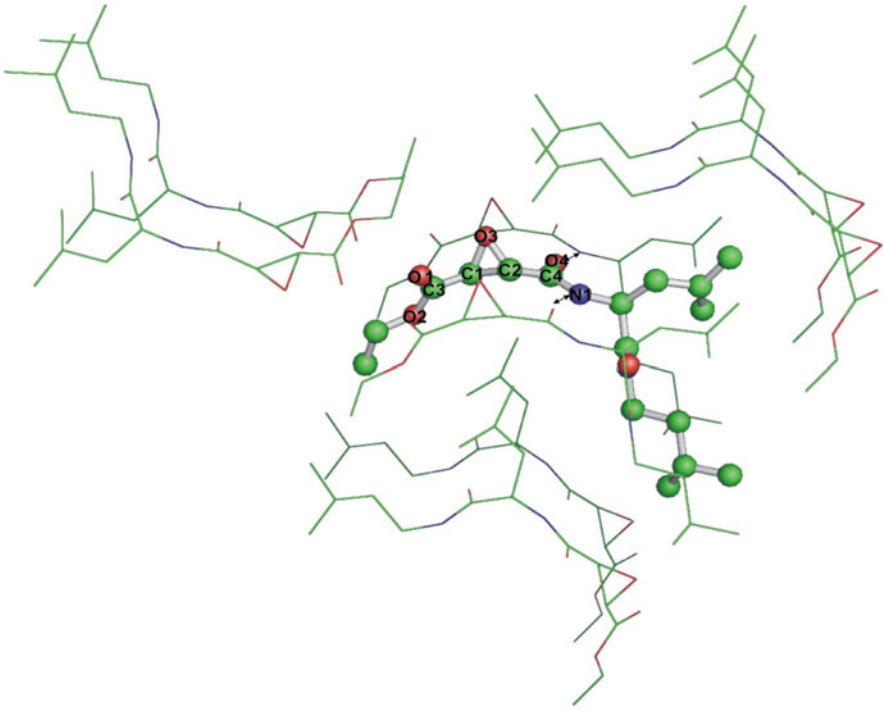


Fig. 16 Orientation of E64d (Loxistatin) in crystals of the pure compound [135]

values at the (3,+1) critical points. In difference to E64c, the C_α carbon atom is the stronger electrophilic center in E64d. This is in line with the experimental result that the α-attack is faster than the β-reaction for ester-substituted epoxides [136, 137]. This changed electrophilicity of C_α and C_β as compared to E64c can be explained by the substitution of the negatively charged carboxylate of E64c which had a strong influence on the epoxide ring by a neutral ester substituent in E64d. Going from gas phase to a polar solvent, the computed value of the Laplacian increases by about 20–25% while an additional rise of 10–15% is obtained for the environment inside of the crystals of E64d (Fig. 17). This shows that also in this case a polar solvent provides a much better approximation of a crystal environment than the gas phase.

For the E64c–enzyme complex, the total effects are much stronger but a considerable part results from the heavily distorted ring structure and the orientation of the carboxylate group. If the geometry taken within the protein–inhibitor complex is also adopted for the polar solvent environment, the Laplacian values obtained for solvent (Fig. 15 p/s) and protein environment [Fig. 15 p/s(S^{-1})], respectively, differ by 20% (β-center) and about 50% (α-center). This shows that the influence of the crystal environment on the β-center of E64d is similar to the influence of the protein

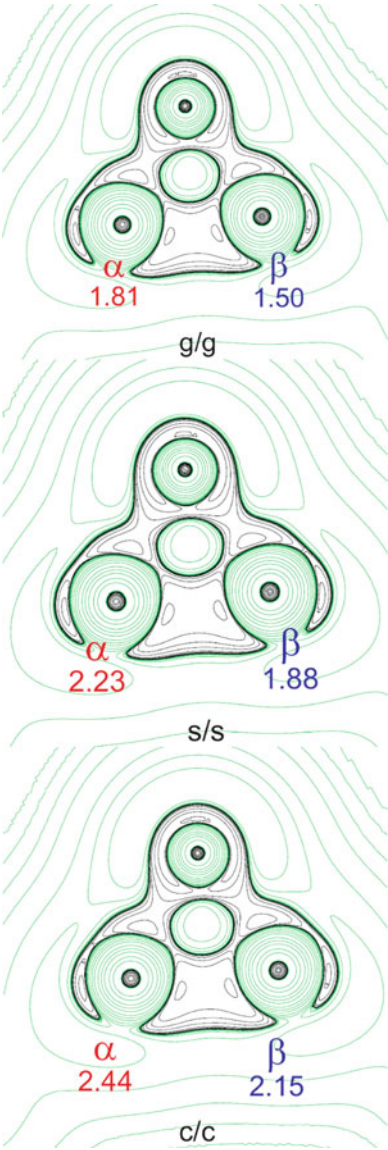


Fig. 17 Contour plot representation of the Laplacian of E64d in the epoxide plane. The values of the (3,+1) critical points are given. Abbreviations for geometry/environment can be taken from Table 16

environment on the β-center of E64c (increase of about 15% as compared to an increase of about 20%). The stronger difference found for the α-centers (increase of about 10% vs. increase of about 50%) is caused by the negatively charged thiolate which is only present in the protein. Such influences in enzymes can obviously not be foreseen from X-ray measurements of the pure compound.

4 On the Applicability of the Source Function in QM/MM Investigations

Bader and Gatti [44, 84] have shown that the ED at any point **r** within a system may be regarded as consisting of contributions from a local source, LS(**r**,**r**′), which operates at all other points of the space:

$$\rho(r) = \int \mathrm{LS}(r, r') \cdot \mathrm{d}r'$$

The local source, given by $LS(r, r') = -(4\pi \cdot |r - r'|)^{-1} \cdot \nabla^2 \rho(r')$, discloses how the cause, the Laplacian of the density at **r**′, gives rise to the effect, the ED at **r**. By integrating the local source over atomic basins Ω defined according to Bader's theory, the ED at **r** may be equated to a sum of contributions S(**r**;Ω), each of which is termed as the *source function contribution* from the atomic basin Ω to ρ(**r**) [84]. A full account of the derivation, properties, and potential uses of the source function approach is given in a separate chapter of this Structure and Bonding volume [44]. The source function represents a measure of how an atom or group of atoms Ω contributes to determine the density at **r**, relative to the contributions from the other atoms or group of atoms in the system. Because of this property, the source function seems to be a convenient instrument to measure influences of the environment on the ED of a biologically active compound within an enzyme. However, it was developed for purely quantum mechanical EDs, or for densities derived from experiment, which, at least in their origin, are also quantum mechanical objects. For that purpose, it must be assured first that the source function delivers reliable results within the QM/MM approach described in the previous chapter. The QM/MM method accounts for the interactions between QM and MM region by an electrostatic embedding scheme [93] which incorporates the MM charges into the one-electron QM Hamiltonian and thereby allows the QM system to adapt to the field exerted by the environment. Dangling bonds at the QM/MM boundary are capped with hydrogen link atoms [138] in the framework of the charge shift method. Point charges do not act as sources for positions other than their own locations and thus do not contribute in a direct manner to the source function, but only indirectly through their effect on the one-electron part of the Hamiltonian and consequently on the ED of the part of the global system modeled through QM. Hence, the source function can only be applied reasonably if all contributions to the density at the investigated point **r** are included in the QM part. The same holds true for the treatment of dangling bonds by the hydrogen link atoms method. To gain information about the necessary size of the QM part for source function applications, we studied several model systems. Hydrocarbon chains are investigated to study the influence of the hydrogen capping treatment. The situation for hydrogen bonding networks described by the electrostatic embedding is examined by water clusters.

As the local source contributions (positive or negative, as a function of the local Laplacian value) have a $(|\mathbf{r}-\mathbf{r}'|)^{-1}$dependence [84], the largest positive and negative contributions to a selected point in the ED should arise from the atoms in the near vicinity and should decrease with larger distances. However, taken separately they possess well-known slow convergence. If they are considered together, a more rapid convergence of atomic SF values with distance from the rp is obtained because of the vanishing of the integrated Laplacian within an atomic basin. Indeed, for very large distances of a basin from the reference point **r**, the importance of the different local weighting of Laplacian values with $(|\mathbf{r}-\mathbf{r}'|)^{-1}$ decreases and at the limit one may replace $(|\mathbf{r}-\mathbf{r}'|)^{-1}$with 1/r which results, in this limiting case, in a null SF contribution (since 1/r may be taken out from the integral and, by definition, the integral of the Laplacian over a QTAIM basin is zero). To test the actual decay, we analyzed the topological parameters of the terminal C–H and C–C bond for the series of hydrocarbons ranging from methane to decane. All structures were chosen in an almost linear zigzag conformation and fully optimized at the B3LYP/TZVP level of theory. The resulting EDs were analyzed for critical points according to the QTAIM theory, and selected BCPs were chosen for the source function analysis. These BCPs were those of a terminal C–H bond and the C–C bond between a terminal methyl group and the adjacent CH_2 unit (see Fig. 18).

The results are summarized in Tables 18a and 19a. The computations showed that the value of the ED at the BCP of the terminal C–H bond BCP_A are almost constant at 1.86 e/Å^3 throughout the complete series. A similar behavior can be found for BCP_B at the terminal C–C bond for which the EDs are found to be 1.64 e/Å^3. The source function analysis shows that for both critical points, the major contributions to the ED arise from the two atoms forming the corresponding bond. Nevertheless, about 12% of the contributions to BCP_A and 23% of those to BCP_B come from the remaining atoms. As expected and especially so for covalently bonded atoms, the contribution of an atom to the ED decreases rapidly with its distance to the considered BCP. Already the contributions of atoms further than two bonds away from the BCP are almost negligible.

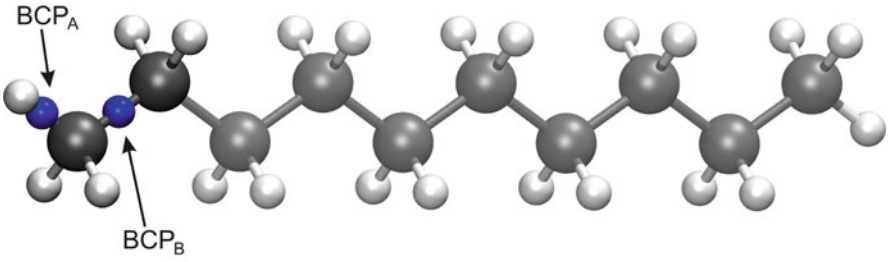


Fig. 18 Structure of the decane molecule and the two bond critical points, one along the terminal C-H bond BCP_A, the other along the bond between the terminal methyl group and the next atom BCP_B

Table 18 Analysis of the ED at the BCP of the terminal CH bond (BCP_A). EDs are given in ($e/Å^3$), Laplacians in ($e/Å^5$), and the contributions in percent

	ρ(BCP)	$\nabla^2\rho$(BCP)	Source function contributions to the ED at the BCP			
			Adjacent[a]	One bond[b]	Two bonds[b]	Three or more bonds[b]
(a) *Structures fully optimized*						
Methane	1.86	−22.39	90	10	–	–
Ethane	1.87	−22.45	89	9	3	–
Propane	1.87	−22.43	89	8	2	1
Butane	1.87	−22.41	89	8	2	1
Pentane	1.87	−22.42	89	8	2	1
Hexane	1.87	−22.41	89	8	2	1
Heptane	1.87	−22.41	89	8	2	1
Octane	1.87	−22.41	89	8	2	1
Nonane	1.87	−22.41	89	8	2	1
Decane	1.86	−22.40	89	8	2	1
(b) *Truncated systems at the geometry of decane, saturated with hydrogen*						
Methane	1.87	−22.79	92	8	–	–
Ethane	1.86	−22.34	89	9	2	–
Propane	1.86	−22.41	89	8	2	1
Butane	1.87	−22.42	89	8	2	1
Pentane	1.87	−22.41	89	8	2	1
Hexane	1.87	−22.41	89	8	2	1
Heptane	1.87	−22.40	89	8	2	1
Octane	1.87	−22.40	89	8	2	1
Nonane	1.86	−22.40	89	8	2	1
Decane	1.86	−22.40	89	8	2	1
(c) *Truncated systems at the geometry of decane saturated with hydrogen, missing atoms replaced by point charges*						
Methane	1.88	−22.99	92	8	–	–
Ethane	1.86	−22.43	89	9	2	–
Propane	1.87	−22.49	89	8	2	1
Butane	1.87	−22.48	89	8	2	1
Pentane	1.87	−22.46	89	8	2	1
Hexane	1.87	−22.44	89	8	2	1
Heptane	1.87	−22.43	89	8	2	1
Octane	1.87	−22.42	89	8	2	1
Nonane	1.87	−22.43	89	8	2	1
Decane	1.86	−22.40	89	8	2	1

[a]Sum of contributions of the atoms which form the bond
[b]Sum of contributions of the atoms being x covalent bonds away

To differentiate between direct effects and the influence of the geometry, we took decane and successively shortened the chain by replacing the terminal methyl group by a hydrogen atom. The molecular structure was not optimized but frozen at the decane geometry. The results are also collected in Tables 18b and 19b. As shown in these tables, the ED at the terminal C–H bond does not change at all. Furthermore, for chains longer than ethane, also the major contributions stay nearly constant. A small variation is only observed between ethane and methane. Due to the substitution of the alkyl chain by a hydrogen atom, the terminal CH bond is artificially elongated, whereby the contribution of this hydrogen atom is lowered by

Table 19 Analysis of the ED at the BCPB of the terminal C–C bond. EDs are given in (e/Å^3), Laplacians in (e/Å^5), and the contributions in percent

	ρ(BCP)	$\nabla^2\rho$(BCP)	Source function contributions to the ED at the BCP			
			Adjacent[a]	One bond[b]	Two bonds[b]	Three or more bonds[b]
(a) *Structures fully optimized*						
Ethane	1.86	−22.40	79	21	–	–
Propane	1.63	−14.29	78	20	3	–
Butane	1.64	−14.38	78	19	2	1
Pentane	1.64	−14.37	78	19	2	1
Hexane	1.64	−14.37	78	19	2	1
Heptane	1.64	−14.38	78	19	2	1
Octane	1.64	−14.39	78	19	2	1
Nonane	1.64	−14.39	78	19	2	1
Decane	1.64	−14.39	79	21	0	0
(b) *Truncated systems at the geometry of decane, saturated with hydrogen*						
Ethane	1.64	−14.41	81	19	–	–
Propane	1.64	−14.35	78	20	2	–
Butane	1.64	−14.41	78	19	2	1
Pentane	1.64	−14.45	78	19	2	1
Hexane	1.64	−14.45	78	19	2	1
Heptane	1.64	−14.45	78	19	2	1
Octane	1.64	−14.45	78	19	2	1
Nonane	1.64	−14.45	78	19	2	1
Decane	1.64	−14.45	78	19	2	1
(c) *Truncated systems at the geometry of decane, saturated with hydrogen, missing atoms replaced by point charges*						
Ethane	1.64	−14.44	81	19	–	–
Propane	1.64	−14.33	78	20	2	–
Butane	1.64	−14.40	78	19	2	1
Pentane	1.64	−14.44	78	19	2	1
Hexane	1.64	−14.45	78	19	2	1
Heptane	1.64	−14.45	78	19	2	1
Octane	1.64	−14.45	78	19	2	1
Nonane	1.64	−14.45	78	19	2	1
Decane	1.64	−14.45	78	19	2	1

[a]Sum of contributions of the atoms which form the bond
[b]Sum of contributions of the atoms being x covalent bonds away

a factor of two. This loss is compensated by an increased contribution arising from the remaining carbon atom, so that the overall density remains constant. A similar trend can be found for the terminal C–C bond (Table 19b) for structures starting from ethane.

As expected, the variations in the contributions are even smaller if the hydrogen-linked atom method is used (Tables 18c and 19c). In summary, the investigations showed that for the present kind of systems and chemical interactions (this conclusion for instance would not necessarily hold for some transition metal complexes) the source function can be used in QM/MM if the second neighbors are included in the QM part. Qualitatively reasonable data are already obtained if only the next neighbors are included in the QM part.

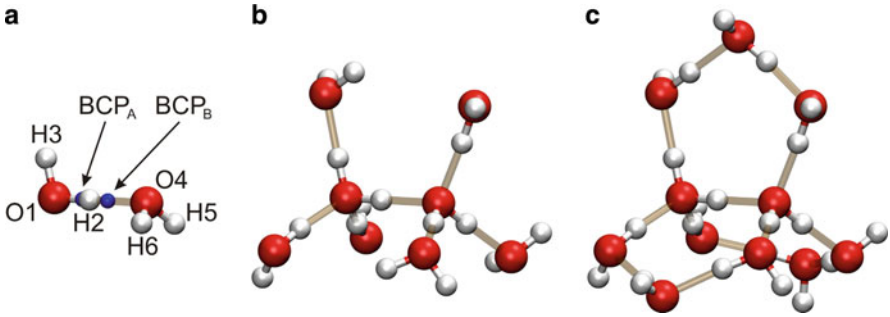


Fig. 19 Geometry of the water clusters under investigation. (**a**) Structure of the central dimer containing the two BCPs of interest. (**b**) Cluster with all surrounding water molecules of the first water shell. (**c**) Optimized structure of the whole $(H_2O)_{11}$ cluster (first and second shell)

The decay of the contributions to the density of a given point in space across hydrogen bonded networks was tested through computations of water clusters (Fig. 19). The common motif of all clusters is the central dimer of two molecules of water forming a hydrogen bond (Fig. 19a). As environment, this dimer is surrounded by up to nine additional water molecules, arranged in such a way that the number of hydrogen bonds toward the central dimer is maximized. Thereby, the central dimer molecules form four hydrogen bonds, clearly more than the average value found for liquid water. Nevertheless, this structure is ideally suited for the intended investigations. The complete system (see Fig. 19c) comprising 11 water molecules was fully optimized at the B3LYP/TZVP level. The system including only the first shell of water molecules around the dimer is shown in Fig. 19b.

Starting from the structure in Fig. 19c, 63 smaller clusters were obtained by replacing more and more of the surrounding molecules. The structures were not further optimized to avoid a bias due to a change in the geometry. To account for the different approaches frequently used for investigations of protein–inhibitor systems, the EDs for all clusters were computed in three different ways. In a first model, the EDs were determined in pure gas phase calculation, only considering the atoms of the given truncated cluster. The second approach uses a polarizable continuum model approximating surrounding liquid water with the standard parameters provided by the GAUSSIAN03 Package [139]. The third system models the surrounding water molecules by means of a QM/MM approach, using a field of point charges to describe the electrostatic influence on the QM system. In these calculations, the deleted atoms in the truncated clusters were replaced by the point charges used in the DL_POLY code [106]. For all systems, the EDs were computed at the B3 [108]-LYP [109, 110]/TZVP [111] level, partitioned according to the QTAIM theory and analyzed with the source function. The analysis was carried out at two BCPs, one along the covalent OH bond between the oxygen atom O1 and the hydrogen atom H2 (BCP_A), the second BCP (BCP_B) along the hydrogen bond linking the two molecules of the central dimer. These BCPs are shown in Fig. 19a.

Table 20 Analysis of the ED at the BCP_A within a single molecule (**a**) and at the BCP_B of the hydrogen bond linking two molecules (**b**)

	ρ(BCP)	$\nabla^2\rho$(BCP)	Source function contributions to the ED at the BCP		
			Dimer	1st shell	2nd shell
(a)					
Dimer	2.07	−46.68	2.07	−46.68	100
Dimer + 1st shell	2.06	−46.45	2.06	−46.45	98
Dimer + 1st and 2nd shell	2.06	−46.46	2.06	−46.46	98
(b)					
Dimer	0.34	3.16	100	–	–
Dimer + 1st shell	0.34	3.04	90	11	0
Dimer + 1st and 2nd shell	0.34	3.01	90	10	2

EDs are given in (e/Å^3), Laplacians in (e/Å^5), and the contributions in percent

The datasets are collected in Table 20. It summarizes the computed densities, Laplacians, and contributions for model 1, i.e., pure gas phase calculations, only considering the atoms of the given cluster. We refrain from giving the values for models 2 (using COSMO) and 3 (using point charges) since the values do not differ significantly from those obtained in model 1. As expected, the EDs at the covalent BCP_A (about 2.09 e/Å^3) are almost six times higher than those at the H-bond BCP_B (about 0.34 e/Å^3). This reflects the different nature of the two bonds. The deviations of these values along the series of clusters are almost negligible ($\pm$0.07 e/Å^3 for BCP_A and less than $\pm$0.01 e/Å^3 for BCP_B), although the surrounding changes dramatically from a complete coverage in the largest clusters to the bare dimer. This is in line with the low sensitivity of the density that is generally observed with respect to environmental changes. Unexpectedly, the same also holds for the Laplacians.

The source function was used to obtain a deeper insight into possible differences of the environmental influences on the ED. As for the hydrocarbon chains, more than 95% of the total contributions to BCP_A originate from the two atoms (O1 and H2) which form the bond. Even the first shell of surrounding atoms contributes less than 5%.

The ED at the H-bonded BCP_B behaves somewhat different. In contrast to the covalent bond, the ED at this point is mainly determined by the heavy atoms in the vicinity, namely O1 and O4 (see Fig. 19a). Each of them contributes roughly 40% to the ED. Although the atom H2 is the nearest atom to the BCP_B and even though it is directly involved in forming the bond, its contribution to the ED is very small, in most cases even negative. This is in agreement with the results of Gatti [47] for the pure dimer, where the corresponding hydrogen atom H2 has also been found to act as a sink in the range of the equilibrium distance.

For the BCP_B of the hydrogen bond, the first shell of the surrounding water molecules contributes about 10% of the ED. Hence, describing the electronic structure in the hydrogen bonding network between enzyme and inhibitor requires only the residues of the protein to be included in the QM part which are directly involved in forming the hydrogen bond.

Using more sophisticated models (replacing the missing water molecules by COSMO or point charges) yields almost identical values. Hence, we refrain from further discussions which would give no additional insights.

5 Summary and Conclusions

This chapter reviews works that examined how different environments affect EDs and related properties. The first part summarized investigations on the reliability of pure theoretical approaches which showed that while absolute numbers for electron densities and Laplacian values are not easily obtained by such theoretical methods, they are suitable for a reasonable reproduction of the trends related to the properties of EDs. The second part considers whether properties of the ED can be used to understand the inhibition potency of possible drugs. The examples of AMCHA, a reversible inhibitor, and of E64c or E64d, as prototypes for irreversible inhibitors, were discussed in detail. AMCHA binds to the target enzyme via intermolecular interactions such as hydrogen bonds, salt bridges, or hydrophobic effects, whereas the latter inhibitors form strong covalent bonds to their target enzyme. In both cases, EDs and properties were computed in four different environments: gas phase, polar solvent, crystal, and inhibitor–enzyme complex. Polar solvents were mimicked by the COSMO method which accounts for the polarizability of the surrounding medium. For crystal and protein surrounding, the atomistic nature of the environment was explicitly represented by a QM/MM approach. To distinguish between direct environmental influences and variations which result more indirectly from different geometries, various model systems were investigated.

In all cases, the influence of the environment on the ED increases in the order gas phase $<$ polar solvent $<$ crystal $<$ protein. This shows again that proteins can polarize molecules stronger than environments inside of crystals of the pure compound. For the reversible inhibitor, AMCHA, the deviations due to different environments are reflected in the computed data, but they are smaller than the uncertainties of high resolution X-ray experiments (0.1 e/Å^3 for EDs, 4–5 e/Å^5 for Laplacians). This supports the assumption that for such inhibitors the ED taken from high-resolution X-ray measurements of crystals of the pure compound can be transferred to related enzyme–inhibitor complexes. It also explains why data taken from the Cambridge structure database are useful for scoring functions used in docking procedures. The differences found for the irreversible inhibitor E64c are much stronger. Hence, the very strong intermolecular bonding interactions between the carboxylate and the enzyme–environment pull the inhibitor onto the thiolate. As a result, the C_α–O bond of the epoxide is considerably weakened and the reactant structure is strongly distorted. The resulting change in the ED is so strong that it cannot be foreseen by measurements or computations in other surroundings. Hence, prediction of the reactivity of irreversible inhibitors

on the basis of high-resolution X-ray measurements of the pure compound might give wrong results.

The computations indicate that gas phase results often deviate quite strongly from the corresponding crystal or enzyme–inhibitor complex data. The agreement improves if the geometrical arrangement of the inhibitor–enzyme complex is used for the gas phase calculations, but the deviations are still considerable. A much better approximation is obtained if the environmental effects are mimicked by the influence of a polar solvent via continuum models. The resulting EDs include polarizations which approximate the situation in crystals or enzyme–inhibitor complexes very well as long as the geometries do not differ too much. Deviations between the less expensive COSMO computation and the QM/MM calculations are often smaller than 1% if the computation is performed for the geometrical structure, which the inhibitor adopts in the crystal of the pure compound or in the inhibitor–enzyme complex. Hence, the former approach would be very useful for databases of average multipole populations or aspherical atomic density functions.

Finally, the Source Function descriptor was scrutinized for its potential to disentangle the effect of the environment on the charge density distribution of the inhibitor species. A preliminary investigation on hydrocarbon linkages and water clusters of increasing size showed that the inclusion of the second neighbors in the QM part allows replicating the effects of the surrounding on the ED features of typical covalent or hydrogen bonded interactions.

References

1. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev B 136(3B): B864–B871
2. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140(4A):1133–1138
3. Coppens P (1997) X-ray charge densities and chemical bonding. Oxford University Press, Oxford
4. Tsirelson VG, Ozerov RP (1996) Electron density and bonding in crystals. IOP Publishing, Bristol
5. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Clarendon, New York
6. Koritsanszky TS, Coppens P (2001) Chemical applications of X-ray charge-density analysis. Chem Rev 101(6):1583–1627. doi:10.1021/Cr990112c
7. Koch W, Holthausen MC (1999) A chemist's guide to density functional theory. Wiley-VCH, Weinheim
8. Ernzerhof M, Perdew JP, Burke K (1996) Density functionals: where do they come from, why do they work? In: Density functional theory I, vol 180. Topics in current chemistry. Springer, Berlin 33, pp 1–30
9. Bader RWF (1990) Atoms in molecules: a quantum theory. Oxford University Press, Oxford
10. Matta CF, Boyd RJ (2007) The quantum theory of atoms in molecules. Wiley-VCH, Weinheim
11. Shaik SS, Hilberty PC (2007) The chemist's guide to valence bond theory. Wiley, Hoboken, NJ

12. Bagus PS, Hermann K, Bauschlicher CW (1984) A new analysis of charge-transfer and polarization for ligand-metal bonding – model studies of Al4co and Al4nh3. J Chem Phys 80 (9):4378–4386
13. Mulliken RS (1955) Electronic population analysis on Lcao-Mo molecular wave functions. 1. J Chem Phys 23(10):1833–1840
14. Reed AE, Curtiss LA, Weinhold F (1988) Intermolecular interactions from a natural bond orbital, donor–acceptor viewpoint. Chem Rev 88(6):899–926
15. Debye P (1915) Zerstreuung von Röntgenstrahlen. Annalen der Physik 351(6):809–823. doi:10.1002/andp.19153510606
16. Coppens P (2005) Charge densities come of age. Angew Chem Int Ed 44(42):6810–6811. doi:10.1002/anie.200501734
17. Leusser D, Henn J, Kocher N, Engels B, Stalke D (2004) S=N versus S+-N-: an experimental and theoretical charge density study. J Am Chem Soc 126(6):1781–1793. doi:10.1021/Ja038941+
18. Kocher N, Henn J, Gostevskii B, Kost D, Kalikhman I, Engels B, Stalke D (2004) Si–E (E = N, O, F) bonding in a hexacoordinated silicon complex: new facts from experimental and theoretical charge density studies. J Am Chem Soc 126(17):5563–5568. doi:10.1021/Ja038459r
19. Hibbs DE, Austin-Woods CJ, Platts JA, Overgaard J, Turner P (2003) Experimental and theoretical charge density study of the neurotransmitter thurine. Chem Eur J 9(5):1075–1084
20. Spackman MA, Munshi P, Dittrich B (2007) Dipole moment enhancement in molecular crystals from X-ray diffraction data. Chemphyschem 8(14):2051–2063. doi:10.1002/cphc.200700339
21. Lecomte C, Souhassou M, Pillet S (2003) Topology of experimental charge density: a tool for understanding atomic interactions. J Mol Struct 647(1–3):53–64. doi:10.1016/s0022-2860(02)00524-0
22. Jelsch C, Guillot B, Lagoutte A, Lecomte C (2005) Advances in protein and small-molecule charge-density refinement methods using MoPro. J Appl Crystallogr 38:38–54. doi:10.1107/S0021889804025518
23. Jelsch C, Teeter MM, Lamzin V, Pichon-Pesme V, Blessing RH, Lecomte C (2000) Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. Proc Natl Acad Sci USA 97(7):3171–3176
24. Luger P (2007) Fast electron density methods in the life sciences – a routine application in the future? Org Biomol Chem 5(16):2529–2540. doi:10.1039/B706235d
25. Cachau R, Howard E, Barth P, Mitschler A, Chevrier B, Lamour V, Joachimiak A, Sanishvili R, Van Zandt M, Sibley E, Moras D, Podjarny A (2000) Model of the catalytic mechanism of human aldose reductase based on quantum chemical calculations. J Phys IV 10(P10):3–13
26. Flaig R, Koritsanszky T, Janczak J, Krane HG, Morgenroth W, Luger P (1999) Fast experiments for charge-density determination: topological analysis and electrostatic potential of the amino acids L-Asn, DL-Glu, DL-Ser, and L-Thr. Angew Chem Int Ed 38(10):1397–1400
27. Li X, Wu G, Abramov YA, Volkov AV, Coppens P (2002) Application of charge density methods to a protein model compound: calculation of Coulombic intermolecular interaction energies from the experimental charge density. Proc Natl Acad Sci USA 99(19):12132–12137. doi:10.1073/pnas.192438999
28. Flaig R, Koritsanszky T, Zobel D, Luger P (1998) Topological analysis of the experimental electron densities of amino acids. 1. D, L-aspartic acid at 20 K. J Am Chem Soc 120 (10):2227–2238
29. Zhurova EA, Zhurov VV, Chopra D, Stash AI, Pinkerton AA (2009) 17 Alpha-estradiol center dot 1/2 H2O: super-structural ordering, electronic properties, chemical bonding, and biological activity in comparison with other estrogens. J Am Chem Soc 131(47): 17260–17269. doi:10.1021/Ja906057z
30. Hansen NK, Coppens P (1978) Electron population analysis of accurate diffraction data. 6. Testing aspherical atom refinements on small-molecule data sets. Acta Crystallogr A 34 (Nov):909–921

31. Volkov A, Abramov Y, Coppens P, Gatti C (2000) On the origin of topological differences between experimental and theoretical crystal charge densities. Acta Crystallogr A 56:332–339
32. Volkov A, Coppens P (2001) Critical examination of the radial functions in the Hansen–Coppens multipole model through topological analysis of primary and refined theoretical densities. Acta Crystallogr A 57:395–405
33. Koritsanszky T, Volkov A, Coppens P (2002) Aspherical-atom scattering factors from molecular wave functions. 1. Transferability and conformation dependence of atomic electron densities of peptides within the multipole formalism. Acta Crystallogr A 58:464–472. doi:10.1107/s0108767302010991
34. Zuo JM, Kim M, O'Keeffe M, Spence JCH (1999) Direct observation of d-orbital holes and Cu–Cu bonding in Cu_2O. Nature 401(6748):49–52
35. Wang SG, Schwarz WHE (2000) Final comment on the discussions of "The case of cuprite". Angew Chem Int Ed 39(21):3794–3796
36. Schwarz WHE (2006) Measuring orbitals: provocation or reality? Angew Chem Int Ed 45 (10):1508–1517. doi:10.1002/anie.200501333
37. Hibbs DE, Overgaard J, Platts JA, Waller MP, Hursthouse MB (2004) Experimental and theoretical charge density studies of tetrafluorophthalonitrile and tetrafluoroisophthalonitrile. J Phys Chem B 108(11):3663–3672. doi:10.1021/Jp0377001
38. Flaig R, Koritsanszky T, Dittrich B, Wagner A, Luger P (2002) Intra- and intermolecular topological properties of amino acids: a comparative study of experimental and theoretical results. J Am Chem Soc 124(13):3407–3417. doi:10.1021/Ja011492y
39. Overgaard J, Waller MP, Platts JA, Hibbs DE (2003) Influence of crystal effects on molecular charge densities in a study of 9-ethynyl-9-fluorenol. J Phys Chem A 107 (50):11201–11208. doi:10.1021/jp036269x
40. Arnold WD, Sanders LK, McMahon MT, Volkov RV, Wu G, Coppens P, Wilson SR, Godbout N, Oldfield E (2000) Experimental, Hartree–Fock, and density functional theory investigations of the charge density, dipole moment, electrostatic potential, and electric field gradients in L-asparagine monohydrate. J Am Chem Soc 122(19):4708–4717
41. Wagner A, Flaig R, Zobel D, Dittrich B, Bombicz P, Strumpel M, Luger P, Koritsanszky T, Krane HG (2002) Structure and charge density of a C-60-fullerene derivative based on a high resolution synchrotron diffraction experiment at 100 K. J Phys Chem A 106(28):6581–6590. doi:10.1021/jp0145199
42. Tafipolsky M, Scherer W, Ofele K, Artus G, Pedersen B, Herrmann WA, McGrady GS (2002) Electron delocalization in acyclic and N-heterocyclic carbenes and their complexes: a combined experimental and theoretical charge-density study. J Am Chem Soc 124(20):5865–5880. doi:10.1021/Ja011761k
43. Ponec R, Gatti C (2009) Do the structural changes defined by the electron density topology necessarily affect the picture of the bonding? Inorg Chem 48(23):11024–11031. doi:10.1021/Ic901197b and references herein
44. Gatti C (2010) The source function descriptor as a tool to extract chemical information from theoretical and experimental electron densities. Structure and bonding. Springer
45. Peres N, Boukhris A, Souhassou M, Gavoille G, Lecomte C (1999) Electron density in ammonium dihydrogen phosphate: non-uniqueness of the multipolar model in simple inorganic structures. Acta Crystallogr A 55:1038–1048
46. Gatti C (2005) Chemical bonding in crystals: new directions. Z Kristallogr 220 (5–6):399–457
47. Bertini L, Cargnoni F, Gatti C (2007) Chemical insight into electron density and wave functions: software developments and applications to crystals, molecular complexes and materials science. Theor Chem Acc 117(5–6):847–884. doi:10.1007/s00214-006-0208-z
48. Podjarny A, Howard E, Mitschler A, Chevrier B, Lecomte C, Guillot B, Pichon-Pesme V, Jelsch C (2002) X-ray crystallography at subatomic resolution. Europhys News 33 (4):113–117
49. Schmidt A, Lamzin VS (2002) Veni, vidi, vici – atomic resolution unravelling the mysteries of protein function. Curr Opin Struct Biol 12(6):698–703

50. Henn J, Ilge D, Leusser D, Stalke D, Engels B (2004) On the accuracy of theoretically and experimentally determined electron densities of polar bonds. J Phys Chem A 108 (43):9442–9452. doi:10.1021/Jp047840a
51. Gatti C, MacDougall PJ, Bader RFW (1988) Effect of electron correlation on the topological properties of molecular charge-distributions. J Chem Phys 88(6):3792–3804
52. Boyd RJ, Ugalde JM (1992) Computational chemistry: structure, interactions and reactivity. Elsevier, Amsterdam
53. Wang J, Shi Z, Boyd RJ, Gonzalez CA (1994) A comparative-study of electron-densities in carbon-monoxide calculated from conventional ab-initio and density-functional methods. J Phys Chem 98(28):6988–6994
54. Boyd RJ, Wang J, Eriksson LA (1995) Recent advances in density functional methods. World Scientific, Singapore
55. Jayatilaka D, Dittrich B (2008) X-ray structure refinement using aspherical atomic density functions obtained from quantum-mechanical calculations. Acta Crystallogr A 64(3):383–393. doi:10.1107/s0108767308005709
56. Cheeseman JR, Carroll MT, Bader RFW (1988) The mechanics of hydrogen-bond formation in conjugated systems. Chem Phys Lett 143(5):450–458
57. Cremer D, Kraka E (1984) A description of the chemical bond in terms of local properties of electron-density and energy. Croat Chem Acta 57(6):1259–1281
58. Ehrlich P (1913) Chemotherapeutics: scientific principles, methods and results. Lancet 182:445–451
59. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci USA 44(2):98–104
60. Koshland DE (1994) The key-lock theory and the induced fit theory. Angew Chem Int Ed 33 (23–24):2375–2378
61. Schmuck C, Engels B, Schirmeister T, Fink R (2008) Chemie fuer Mediziner. Pearson, Muenchen
62. Berg JM, Tymoczko JL, Stryer L (2007) Biochemistry. WH Freeman, New York
63. Schmidt A, Lamzin VS (2007) From atoms to proteins. Cell Mol Life Sci 64(15):1959–1969. doi:10.1007/s00018-007-7195-7
64. Cachau RE, Podjarny AD (2005) High-resolution crystallography and drug design. J Mol Recognit 18(3):196–202. doi:Doi 10.1002/Jmr.738
65. Howard EI, Sanishvili R, Cachau RE, Mitschler A, Chevrier B, Barth P, Lamour V, Van Zandt M, Sibley E, Bon C, Moras D, Schneider TR, Joachimiak A, Podjarny A (2004) Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 angstrom. Proteins 55(4):792–804
66. Muzet N, Guillot B, Jelsch C, Howard E, Lecomte C (2003) Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations. Proc Natl Acad Sci USA 100(15):8742–8747. doi:10.1073/pnas.1432955100
67. Lamour V, Barth P, Rogniaux H, Poterszman A, Howard E, Mitschler A, Van Dorsselaer A, Podjarny A, Motas D (1999) Production of crystals of human aldose reductase with very high resolution diffraction. Acta Crystallogr D 55:721–723
68. Grabowsky S, Pfeuffer T, Morgenroth W, Paulmann C, Schirmeister T, Luger P (2008) A comparative study on the experimentally derived electron densities of three protease inhibitor model compounds. Org Biomol Chem 6(13):2295–2307. doi:10.1039/B802831a
69. Grabowsky S, Pfeuffer T, Checinska L, Weber M, Morgenroth W, Luger P, Schirmeister T (2007) Electron-density determination of electrophilic building blocks as model compounds for protease inhibitors. Eur J Org Chem (17):2759–2768. doi:10.1002/ejoc.200601074
70. Ghermani NE, Spasojevic-de Bire A, Bouhmaida N, Ouharzoune S, Bouligand J, Layre A, Gref R, Couvreur P (2004) Molecular reactivity of busulfan through its experimental electrostatic properties in the solid state. Pharm Res 21(4):598–607

71. Wagner A, Flaig R, Dittrich B, Schmidt H, Koritsanszky T, Luger P (2004) Charge density and experimental electrostatic potentials of two penicillin derivatives. Chem Eur J 10 (12):2977–2982. doi:10.1002/chem.200305627
72. Flaig R, Koritsanszky T, Soyka R, Haming L, Luger P (2001) Electronic insight into an antithrombotic agent by high-resolution X-ray crystallography. Angew Chem Int Ed 40 (2):355–359
73. Klebe G (1994) The use of composite crystal-field environments in molecular recognition and the de-novo design of protein ligands. J Mol Biol 237(2):212–235
74. Klebe G (2008) Structure correlation and ligand/receptor interactions. Structure correlation. Wiley-VCH Verlag GmbH, Weinheim
75. Velec HFG, Gohlke H, Klebe G (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. J Med Chem 48(20):6296–6303. doi:10.1021/Jm050436v
76. Boer DR, Kroon J, Cole JC, Smith B, Verdonk ML (2001) SuperStar: comparison of CSD and PDB-based interaction fields as a basis for the prediction of protein–ligand interactions. J Mol Biol 312(1):275–287
77. Bruno IJ, Cole JC, Lommerse JPM, Rowland RS, Taylor R, Verdonk ML (1997) IsoStar: a library of information about nonbonded interactions. J Comput Aided Mol Des 11 (6):525–537
78. Dittrich B, Koritsanszky T, Luger P (2004) A simple approach to nonspherical electron densities by using invarioms. Angew Chem Int Ed 43(20):2718–2721. doi:10.1002/anie.200353596
79. Hubschle CB, Dittrich B, Grabowsky S, Messerschmidt M, Luger P (2008) Comparative experimental electron density and electron localization function study of thymidine based on 20 K X-ray diffraction data. Acta Crystallogr B 64:363–374. doi:10.1107/s0108768108005776
80. Volkov A, Koritsanszky T, Li X, Coppens P (2004) Response to the paper. A comparison between experimental and theoretical aspherical-atom scattering factors for charge-density refinement of large molecules, by Pichon-Pesme, Jelsch, Guillot & Lecomte (2004). Acta Crystallogr A 60:638–639. doi:10.1107/S0108767304016496
81. Dittrich B, Weber M, Kalinowski R, Grabowsky S, Hubschle CB, Luger P (2009) How to easily replace the independent atom model – the example of bergenin, a potential anti-HIV agent of traditional Asian medicine. Acta Crystallogr B 65:749–756. doi:10.1107/S0108768109046060
82. Dittrich B, Hubschle CB, Holstein JJ, Fabbiani FPA (2009) Towards extracting the charge density from normal-resolution data. J Appl Crystallogr 42:1110–1121. doi:10.1107/S0021889809034621
83. Mladenovic M, Arnone M, Fink RF, Engels B (2009) Environmental effects on charge densities of biologically active molecules: do molecule crystal environments indeed approximate protein surroundings? J Phys Chem B 113(15):5072–5082. doi:10.1021/Jp809537v
84. Bader RFW, Gatti C (1998) A Green's function for the density. Chem Phys Lett 287 (3–4):233–238
85. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Salvador P, Dannenberg JJ, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian98. Gaussian, Inc., Pittsburgh, PA
86. Gatti C, Bianchi R, Destro R, Merati F (1992) Experimental vs theoretical topological properties of charge-density distributions – an application to the L-alanine molecule studied by X-ray-diffraction at 23-K. J Mol Struct Theochem 87:409–433

87. Senn HM, Thiel W (2007) QM/MM methods for biological systems. In: Atomistic approaches in modern biology: from quantum chemistry to molecular simulations, vol 268, pp 173–290. doi:10.1007/128_2006_084
88. Monard G, Merz KM (1999) Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. Acc Chem Res 32(10):904–911
89. Gao JL, Truhlar DG (2002) Quantum mechanical methods for enzyme kinetics. Annu Rev Phys Chem 53:467–505. doi:10.1146/annurev.physchem.53.091301.150114
90. Field MJ (2002) Simulating enzyme reactions: challenges and perspectives. J Comput Chem 23(1):48–58
91. Monard G, Prat-Resina X, Gonzalez-Lafont A, Lluch JM (2003) Determination of enzymatic reaction pathways using QM/MM methods. Int J Quantum Chem 93(3):229–244. doi:10.1002/Qua.10555
92. Ridder L, Mulholland AJ (2003) Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: from structure to activity. Curr Top Med Chem 3(11):1241–1256
93. Bakowies D, Thiel W (1996) Hybrid models for combined quantum mechanical and molecular mechanical approaches. J Phys Chem 100(25):10580–10594
94. Field MJ, Bash PA, Karplus M (1990) A combined quantum-mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem 11(6):700–733
95. Amara P, Field MJ (2003) Evaluation of an ab initio quantum mechanical/molecular mechanical hybrid-potential link-atom method. Theor Chem Acc 109(1):43–52. doi:10.1007/s00214-002-0413-3
96. Reuter N, Dejaegere A, Maigret B, Karplus M (2000) Frontier bonds in QM/MM methods: a comparison of different approaches. J Phys Chem A 104(8):1720–1735
97. Singh UC, Kollman PA (1986) A combined abinitio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular systems – applications to the $CH^3Cl + Cl$ – exchange reaction and gas-phase protonation of polyethers. J Comput Chem 7(6):718–730
98. Derat E, Bouquant J, Humbel S (2003) On the link atom distance in the ONIOM scheme. An harmonic approximation analysis. J Mol Struct Theochem 632:61–69. doi:10.1016/s0166-1280(03)00288-4
99. Klamt A, Schuurmann G (1993) Cosmo – a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2(5):799–805
100. Schafer A, Klamt A, Sattel D, Lohrenz JCW, Eckert F (2000) COSMO implementation in TURBOMOLE: extension of an efficient quantum chemical code towards liquid systems. Phys Chem Chem Phys 2(10):2187–2193
101. Mathews II, VanderhoffHanaver P, Castellino FJ, Tulinsky A (1996) Crystal structures of the recombinant kringle 1 domain of human plasminogen in complexes with the ligands epsilon-aminocaproic acid and trans-4-(aminomethyl)cyclohexane-1-carboxylic acid. Biochemistry 35(8):2567–2576
102. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) Charmm – a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4(2):187–217
103. MacKerell ADJ, Brooks BR, Brooks CL III, Nilsson L, Roux B, Won Y, Karplus M (1998) The encyclopedia of computational chemistry, vol 1. Wiley, Chichester
104. Huet G, Flipo RM, Richet C, Thiebaut C, Demeyer D, Balduyck M, Duquesnoy B, Degand P (1992) Measurement of elastase and cysteine proteinases in synovial-fluid of patients with rheumatoid-arthritis, seronegative spondylarthropathies, and osteoarthritis. Clin Chem 38 (9):1694–1697
105. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ, Billeter S, Terstegen F, Thiel S, Kendrick J, Rogers SC, Casci J, Watson M, King F, Karlsen E, Sjovoll M, Fahmi A, Schafer A, Lennartz C (2003) QUASI:

a general purpose implementation of the QM/MM approach and its application to problems in catalysis. J Mol Struct Theochem 632:1–28. doi:10.1016/s0166-1280(03)00285-9
106. Smith W, Forester TR (1996) DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. J Mol Graph 14(3):136–141
107. TURBOMOLE V5.6 2005, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2005. Available from http://www.turbomole.com
108. Becke AD (1993) Density-functional thermochemistry. 3. The role of exact exchange. J Chem Phys 98(7):5648–5652
109. Becke AD (1988) Density-functional exchange-energy approximation with corrected asymptotic behaviour. Phys Rev A 38(6):3098–3100
110. Lee CT, Yang WT, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. Phys Rev B 37(2):785–789
111. Schafer A, Huber C, Ahlrichs R (1994) Fully optimized contracted Gaussian-basis sets of triple zeta valence quality for atoms Li to Kr. J Chem Phys 100(8):5829–5835
112. Schlund S, Muller R, Grassmann C, Engels B (2008) Conformational analysis of arginine in gas phase – a strategy for scanning the potential energy surface effectively. J Comput Chem 29(3):407–415. doi:10.1002/Jcc.20798
113. Schlund S, Mladenovic M, Janke EMB, Engels B, Weisz K (2005) Geometry and cooperativity effects in adenosine–carboxylic acid complexes. J Am Chem Soc 127(46):16151–16158. doi:10.1021/Ja0531430
114. Schlund S, Schmuck C, Engels B (2007) How important is molecular rigidity for the complex stability of artificial host–guest systems? A theoretical study on self-assembly of gas-phase arginine. Chem Eur J 13(23):6644–6653. doi:10.1002/chem.200601741
115. Hupp T, Sturm C, Janke EMB, Cabre MP, Weisz K, Engels B (2005) A combined computational and experimental study of the hydrogen-bonded dimers of xanthine and hypoxanthine. J Phys Chem A 109(8):1703–1712. doi:10.1021/Jp0460588
116. Schlund S, Schmuck C, Engels B (2005) "Knock-out" analogues as a tool to quantify supramolecular processes: a theoretical study of molecular interactions in guanidiniocarbonyl pyrrole carboxylate dimers. J Am Chem Soc 127(31):11115–11124. doi:10.1021/Ja052536w
117. Zhao HM, Pfister J, Settels V, Renz M, Kaupp M, Dehm VC, Wurthner F, Fink RF, Engels B (2009) Understanding ground- and excited-state properties of perylene tetracarboxylic acid bisimide crystals by means of quantum chemical computations. J Am Chem Soc 131 (43):15660–15668. doi:10.1021/Ja902512e
118. Schlund S, Janke EMB, Weisz K, Engels B (2010) Predicting the tautomeric equilibrium of acetylacetone in solution. I. The right answer for the wrong reason? J Comput Chem 31 (4):665–670. doi:10.1002/Jcc.21354
119. Musch PW, Engels B (2001) The importance of the Ene reaction for the C-2-C-6 cyclization of enyne-allenes. J Am Chem Soc 123(23):5557–5562. doi:10.1021/Ja010346p
120. Suter HU, Pless V, Ernzerhof M, Engels B (1994) Difficulties in the calculation of electron-spin-resonance parameters using density-functional methods. Chem Phys Lett 230 (4–5):398–404
121. Groth P (1968) Crystal structure of trans form of 1,4-aminomethylcyclohexanecarboxylic acid. Acta Chem Scand 22(1):143–158
122. Gatti C, Saunders VR, Roetti C (1994) Crystal-field effects on the topological properties of the electron-density in molecular-crystals – the case of urea. J Chem Phys 101(12):10686–10696
123. Roby KR (1974) Quantum-theory of chemical valence concepts. 1. Definition of charge on an atom in a molecule and of occupation numbers for electron-density shared between atoms. Mol Phys 27(1):81–104
124. Heinzmann R, Ahlrichs R (1976) Population analysis based on occupation numbers of modified atomic orbitals (maos). Theor Chim Acta 42(1):33–45
125. Lecaille F, Kaleta J, Bromme D (2002) Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design. Chem Rev 102(12):4459–4488. doi:10.1021/cr0101656

126. Sloane BF, Moin K, Krepela E, Rozhin J (1990) Cathepsin-B and its endogenous inhibitors – the role in tumor malignancy. Cancer Metastasis Rev 9(4):333–352
127. Otto HH, Schirmeister T (1997) Cysteine proteases and their inhibitors. Chem Rev 97 (1):133–171
128. Powers JC, Asgian JL, Ekici OD, James KE (2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. Chem Rev 102(12):4639–4750. doi:10.1021/cr010182v
129. Yamamoto A, Tomoo K, Matsugi K, Hara T, In Y, Murata M, Kitamura K, Ishida T (2002) Structural basis for development of cathepsin B-specific noncovalent-type inhibitor: crystal structure of cathepsin B-E64c complex. Biochim Biophys Acta 1597(2):244–251
130. Vahtras O, Almlof J, Feyereisen MW (1993) Integral approximations for LCAO-SCF calculations. Chem Phys Lett 213(5–6):514–518
131. Mladenovic M, Junold K, Fink RF, Thiel W, Schirmeister T, Engels B (2008) Atomistic insights into the inhibition of cysteine proteases: first QM/MM calculations clarifying the regiospecificity and the inhibition potency of epoxide- and aziridine-based inhibitors. J Phys Chem B 112(17):5458–5469. doi:10.1021/Jp711287c
132. Mladenovic M, Ansorg K, Fink RF, Thiel W, Schirmeister T, Engels B (2008) Atomistic insights into the inhibition of cysteine proteases: first QM/MM calculations clarifying the stereoselectivity of epoxide-based inhibitors. J Phys Chem B 112(37):11798–11808. doi:10.1021/Jp803895f
133. Mladenovic M, Schirmeister T, Thiel S, Thiel W, Engels B (2007) The importance of the active site histidine for the activity of epoxide- or aziridine-based inhibitors of cysteine proteases. Chemmedchem 2(1):120–128. doi:10.1002/cmdc.200600159
134. Mladenovic M, Fink RF, Thiel W, Schirmeister T, Engels B (2008) On the origin of the stabilization of the zwitterionic resting state of cysteine proteases: a theoretical study. J Am Chem Soc 130(27):8696–8705. doi:10.1021/ja711043x
135. Ishida T, Sakaguchi M, Yamamoto D, Inoue M, Kitamura K, Hanada K, Sadatome T (1988) Conformation of Ethyl (+)-(2S, 3S)-3-(1-N-(3-methylbutyl)amino leucyl-carbonyl)oxirane-2-carboxylate (Loxistatin), a cysteine protease inhibitor – X-ray crystallographic and H-1 nuclear magnetic resonance studies. J Chem Soc Perkin Trans 2(6):851–857
136. Helten H, Schirmeister T, Engels B (2005) Theoretical studies about the influence of different ring substituents on the nucleophilic ring opening of three-membered heterocycles and possible implications for the mechanisms of cysteine protease inhibitors. J Org Chem 70 (1):233–237. doi:10.1021/Jo048373w
137. Helten H, Schirmeister T, Engels B (2004) Model calculations about the influence of protic environments on the alkylation step of epoxide, aziridine, and thiirane based cysteine protease inhibitors. J Phys Chem A 108(38):7691–7701. doi:10.1021/Jp048784g
138. de Vries AH, Sherwood P, Collins SJ, Rigby AM, Rigutto M, Kramer GJ (1999) Zeolite structure and reactivity by combined quantum-chemical-classical calculations. J Phys Chem B 103(29):6133–6141
139. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Rob MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03. Gaussian, Inc., Wallingford, CT

Struct Bond (2012) 147: 99–142
DOI: 10.1007/430_2010_37
© Springer-Verlag Berlin Heidelberg 2011
Published online: 8 February 2011

Electron Density in Quantum Theory

Samuel Fux and Markus Reiher

Abstract In this work, we review the theory of the electron density in quantum chemistry and discuss to which extent relativistic effects are recovered by approximate relativistic Hamiltonians. For this purpose, we give an overview on different approximations to the fully relativistic many-electron Hamiltonian. In addition, we present new results, considering correlation effects on the electron density of a transition metal complex.

Keywords Continuity equation · Contact density · Conceptual theories · Dirac–Coulomb–Breit · Density functional theory · Douglas–Kroll–Hess · Electron density · Electron correlation effects · Frozen-density embedding · Picture change error · Relativistic effects · ZORA

Contents

1 Introduction 100
2 Definition of the Electron Density Distribution 102
3 Dependence on Wave Function and Hamiltonian 103
3.1 The Wave Function 103
3.2 The Hamiltonian Operator 106
3.3 Electron Density Distribution for Different Choices of Hamiltonian and Wave Function 112
3.4 Methodological Aspects: Density Functional Theory and Current-Density Functional Theory 115
4 Analysis of Approximate Electron Densities 116
4.1 Effect of the Approximate Hamiltonians 117
4.2 Significance of Electron Correlation: $Fe(NO)^{2+}$ as an Example 120
5 The Most Difficult Case: Contact Densities 125
6 Electron Density in Conceptual Theories 128
6.1 Conceptual Density Functional Theory 128

S. Fux and M. Reiher (✉)
Laboratorium für Physikalische Chemie, ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland
e-mail: markus.reiher@phys.chem.ethz.ch

6.2 The Quantum Theory of Atoms in Molecules 131
7 Electron Density in Embedding Schemes 132
8 Conclusion 136
9 Computational Methodology 137
References 137

1 Introduction

Electron densities obtained from quantum chemical calculations and from experiments, especially from X-ray diffraction, are a powerful quantity for structure and reactivity studies in chemistry. Topological analyses of these electron densities and their negative Laplacian provide a deeper understanding of bonding in molecules and complexes [1–5]. The importance of X-ray diffraction experiments for structure determination is obvious from its early history. The discovery of X-ray diffraction in crystals by von Laue et al. [6] was followed by Bragg's idea [7] that one may determine the arrangement of the atoms in a crystal with this new technique. Once the tremendous potential of X-ray diffraction experiments for structure determination was recognized, also organic compounds (urotropine [8]) as well as biological systems like proteins (e.g., sperm whale myoglobin [9]) were investigated in X-ray diffraction experiments. Experimental electron densities are determined from a diffraction pattern from which one can calculate generalized structure factors [10] that depend on a scattering vector and are described by an intensity amplitude and a phase factor. The amplitudes of the phase factors are obtained from experiment, whereas the phase information is missing. This "phase problem" was solved by Hauptman and Karle [11, 12], who developed direct methods for the calculation of the phase factors. For systems that contain heavy nuclei, there exists also the so-called Patterson method [13, 14]. In practical applications, one usually does not use these Fourier transform methods but a different approach, in which structure factors are first estimated from a guess density for the sought-for structure which is constructed from the atomic densities. The charge density and the structural parameters are then adjusted in a fitting procedure to the experimentally obtained parameters, using either the method of least squares or entropy maximization [15].

In theory, the many-electron wave function $\Psi_n(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N,t)$, with $\boldsymbol{r}_i$ denoting a vector containing the coordinates of electron i and n being the index of the electronic state (which we, however, skip in the following for the sake of brevity), is the central object for the calculation of molecular properties in quantum chemistry. Hence, also the electron density of any molecule or solid can be calculated from this wave function. In principle, a system containing a given number of electrons and atomic nuclei is fully determined by its wave function, which has no physical meaning in the sense that it is not an observable. On the other hand, the external potential (and therefore the energy) of a system is – owing to the first Hohenberg–Kohn theorem [16] – fully determined by its ground-state electron density $\rho_0(\boldsymbol{r})$ which is an observable available in theory and experiment.

The Hohenberg–Kohn theorems are the foundation for the development of density functional theory (DFT) [17] and its extensions, e.g., time-dependent DFT, conceptual DFT, current DFT, and subsystem DFT. These methods are based on the electron density rather than on the wave function as central quantity for the calculation of molecular properties and chemical descriptors. All contributions to the total energy are represented by functionals which depend on the electron density only such that the total energy and in principle even the wave function $\Psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N, t) = \Psi[\rho(\boldsymbol{r}, t)]$ are given as functionals of the density. Early attempts to define such density functionals were made by Thomas [18] and by Fermi [19] presenting the first kinetic energy density functionals in 1927/1928. Molecular properties and chemical descriptors are then defined as derivatives of the total-energy density functional $E[\rho]$ with respect to external perturbations.

In practice, the electron density is usually calculated from a wave function (even within DFT → Kohn–Sham DFT). For this, one has to choose a suitable approximate Hamiltonian operator and an ansatz for the wave function. In order to arrive at a consistent theory that overcomes all pitfalls and covers all interactions and effects important for the chemistry of the whole periodic table, including heavy atoms, one must apply a theory which is based on the Dirac equation [20, 21]. A comprehensive description of matter is therefore solely given by the Dirac–Coulomb–Breit or the Dirac–Coulomb Hamiltonian [22] which we use as the standard reference when we investigate the performance or accuracy of any approximate relativistic method (the electron–electron interaction is usually approximated by the instantaneous Coulomb interaction). The most important approximate Hamiltonian operators will be discussed in detail in the theory section of this work.

Different types of many-electron functions are known as approximations to the exact wave function and are built from one-electron functions, i.e., from orbitals $\psi_i(\boldsymbol{r})$. Such an independent-particle model in which the wave function can be assembled from an antisymmetrized product of N one-electron functions entirely neglects the correlated motion of the electrons and causes therefore errors in the description of systems containing more than one electron. It is therefore important to carry out a systematic analysis of the method-inherent approximations to ensure that a sufficiently high accuracy for electron densities obtained from quantum chemical calculations can be guaranteed (correlation effects on the electron density will be discussed in Sect. 4.2 of this work).

This work is organized as follows. In Sect. 2, we demonstrate how equations for the electron density are derived from fundamental principles to analyze in the next section how these equations depend on the choice of the Hamiltonian and the ansatz for the wave function. Then we proceed in Sect. 4 with the analysis of approximate electron densities obtained from different choices for the Hamiltonian with the focus on relativistic effects. In Sect. 5, we regard the electron density at the position of the nucleus, which is prone to errors for most of the approximate Hamiltonians. Section 6 deals with the electron density in the context of conceptual DFT and atoms in molecules in combination with relativistic electronic structure methods.

In the last section, we have a closer look on a particular branch of DFT, namely frozen-density embedding (FDE) which is a subsystem formulation of DFT that can, for instance, be used to describe a molecule embedded in a crystal, hence allowing for crystal-packing effects on the electron density.

2 Definition of the Electron Density Distribution

For the simplest case, i.e., for a system containing only one electron, the square of the absolute value of the wave function can – according to Born [23] – be interpreted as a probability density distribution from which one can determine the probability of finding the electron at a given infinitesimally small volume at position $\boldsymbol{r} = (x,y,z)$ in space. In order to determine the probability in a finite volume, integration is necessary. The probability of finding the electron anywhere in space must be equal to one at any time:

$$\int_{-\infty}^{+\infty} \mathrm{d}^3 r |\Psi(\boldsymbol{r},t)|^2 = 1. \tag{1}$$

Born's interpretation of the wave function of a single electron can be generalized to an N-electron system, described by the normalized wave function $\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_\mathrm{N},t) = \Psi(\{\boldsymbol{r}_i\},t)$:

$$\rho(\boldsymbol{r}) \stackrel{!}{=} N \int_{-\infty}^{+\infty} \mathrm{d}^3 r_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}^3 r_N |\Psi(\{\boldsymbol{r}_i\},t)|^2 \tag{2}$$

(note that spin coordinates are not considered, at this stage because these somewhat artificial coordinates are not present in relativistic theory; see below). Quantum mechanics states that for any observable a corresponding operator exists, which yields a set of eigenvalues being the possible results of a measurement. Therefore, an operator can also be assigned to the electron density. The expectation value of this density operator

$$\rho_{\boldsymbol{r}} = \sum_{i=1}^{N} \delta^{(3)}(\boldsymbol{r} - \boldsymbol{r}_i) \tag{3}$$

yields the particle density at a given position $\boldsymbol{r}$. Here, $\delta^{(3)}(\boldsymbol{r} - \boldsymbol{r}_i) = \delta(x - x_i)\delta(y - y_i)\delta(z - z_i)$ denotes the three-dimensional delta distribution. The electron density can then be expressed as an expectation value of this density operator

$$\begin{aligned} \rho(\boldsymbol{r}) &\stackrel{!}{=} \langle \Psi(\{\boldsymbol{r}_i\},t) | \rho_{\boldsymbol{r}} | \Psi(\{\boldsymbol{r}_i\},t) \rangle \\ &= \int_{-\infty}^{+\infty} \mathrm{d}^3 r_1 \int_{-\infty}^{+\infty} \mathrm{d}^3 r_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}^3 r_N \, \Psi^*(\{\boldsymbol{r}_i\},t) \rho_{\boldsymbol{r}} \, \Psi(\{\boldsymbol{r}_i\},t), \end{aligned} \tag{4}$$

where in contrast to the first definition of the electron density the integration is now carried out over all dynamical variables. The charge density ρ_c is directly related to the electron density via the negative elementary charge: $\rho_c(\boldsymbol{r}) = -e\rho(\boldsymbol{r})$.

The propagation of this expectation value in time can then be described by the Ehrenfest theorem, which allows us to express the total time derivative of an expectation value as an expectation value of the partial time derivative of the operator and an expectation value of the commutator of the operator with the Hamiltonian:

$$\begin{aligned}\frac{\mathrm{d}\rho(\boldsymbol{r},t)}{\mathrm{d}t} &= \frac{\mathrm{d}\langle\Psi(\{\boldsymbol{r}_i\},t)|\rho_{\boldsymbol{r}}|\Psi(\{\boldsymbol{r}_i\},t)\rangle}{\mathrm{d}t} \\ &= \underbrace{\left\langle \Psi(\{\boldsymbol{r}_i\},t)\left|\frac{\partial\rho_{\boldsymbol{r}}}{\partial t}\right|\Psi(\{\boldsymbol{r}_i\},t)\right\rangle}_{=0} + \underbrace{\frac{i}{\hbar}\langle\Psi(\{\boldsymbol{r}_i\},t)|[H,\rho_{\boldsymbol{r}}]|\Psi(\{\boldsymbol{r}_i\},t)\rangle}_{-\nabla\cdot\boldsymbol{j}}. \quad (5)\end{aligned}$$

The partial derivative of the density operator with respect to time vanishes in this equation, because this operator does not depend on time. If one chooses a Hamiltonian and a wave function, the second term can be evaluated and yields the negative divergence of the current density. Because the position $\boldsymbol{r}$ does not depend on the time t, the total derivative in (5) is equal to the partial derivative $\mathrm{d}\rho(\boldsymbol{r},t)/\mathrm{d}t = \partial\rho(\boldsymbol{r},t)/\partial t$, and one therefore arrives at the continuity equation

$$\frac{\partial\rho(\boldsymbol{r},t)}{\partial t} + \nabla\cdot\boldsymbol{j} = 0, \quad (6)$$

which is the fundamental equation defining both the electron density and the current density $\boldsymbol{j}$ [22]. The deduction of the continuity equation from the expectation value of the density operator $\langle\Psi(\{\boldsymbol{r}_i\},t)|\rho_{\boldsymbol{r}}|\Psi(\{\boldsymbol{r}_i\},t)\rangle$ uniquely defines the density distribution of an N-particle system. The continuity equation can alternatively be deduced from the Heisenberg equation of motion written for the density operator [22], which is omitted here for the sake of brevity.

3 Dependence on Wave Function and Hamiltonian

From the continuity equation, it is clear that one has to choose a wave function and a Hamiltonian operator [see (5)] to resolve the electron density and the current density. We will therefore give a brief overview on approximations to the electronic wave function and on the different Hamiltonians relevant to chemistry before proceeding with the analysis of the electron density.

3.1 The Wave Function

The most general ansatz for the total wave function of a molecule consisting of N electrons and M nuclei,

$$\Psi(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N,\boldsymbol{R}_1,\ldots\boldsymbol{R}_M,t)=\Psi(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\},t), \tag{7}$$

depends on all nuclear coordinates, the coordinates of the electrons, and time (if an absolute time frame is assumed). Since we are interested in the calculation of the stationary electron density, we look for a time-independent wave function that depends only parametrically on the nuclear coordinates. After the separation of time (by a product ansatz) and of the nuclear coordinates (Born–Oppenheimer approximation [24–26]), one arrives at the time-independent electronic wave function $\Psi_m(\{\boldsymbol{r}_i\})$ for electronic state m, which has to be approximated. A simple product ansatz (Hartree product) of one-electron functions (orbitals) violates the Pauli exclusion principle, because the wave function is no longer antisymmetric with respect to the exchange of any two electronic coordinates. A corrected ansatz explicitly implements the Pauli exclusion principle and can also be expressed as a normalized determinant of a set of all N occupied orbitals,

$$\Theta_k(\{\boldsymbol{r}_i\})=\hat{A}\prod_{i=1}^{N}\psi_{k_i}(\boldsymbol{r}_i)=\frac{1}{\sqrt{N!}}\begin{vmatrix}\psi_{k_1}(\boldsymbol{r}_1) & \cdots & \psi_{k_1}(\boldsymbol{r}_N)\\ \vdots & \ddots & \vdots\\ \psi_{k_N}(\boldsymbol{r}_1) & \cdots & \psi_{k_N}(\boldsymbol{r}_N)\end{vmatrix}, \tag{8}$$

with $\hat{A}$ denoting the antisymmetrization operator given by

$$\hat{A}=\frac{1}{\sqrt{N!}}\left(1-\sum_{ij}P_{ij}+\sum_{ijk}P_{ijk}-\cdots\right), \tag{9}$$

where P_{ij} stands for all permutations of the two electrons i and j, P_{ijk} for all possible permutations of the electrons i, j, and k, and so forth. The so-called Slater determinant Θ_k contains either the N orbitals with the lowest energy (corresponding to the ground state in a one-determinant picture) or orbitals with higher orbital energy. In practical applications, a basis set is introduced to represent the one-electron functions.

In general, linear combinations of Slater determinants can be set up to yield eigenfunctions of the squared spin operator. These linear combinations are called configuration state functions (CSFs)

$$\Phi_l=\sum_k b_{lk}\Theta_k, \tag{10}$$

with $\{b_{lk}\}$ being known expansion coefficients. The single-determinant or single-CSF ansatz can be improved by subsequently adding more and more CSFs. If all (infinitely many) possible determinants are considered in the linear combination, one obtains the so-called full configuration interaction (FCI) wave function

$$\Psi_m^{\mathrm{FCI}}(\{\boldsymbol{r}_i\})=\sum_l c_{lm}\Phi_l(\{\boldsymbol{r}_i\})=\sum_l\sum_k c_{lm}b_{lk}\left(\hat{A}\prod_{i=1}^{N}\psi_{k_i}(\boldsymbol{r}_i)\right), \tag{11}$$

with $\{c_{lm}\}$ denoting the CI expansion coefficients of the CSFs. With such a complete many-electron basis set of CSFs, any many-electron function can be expressed as a linear combination of these many-electron basis functions. The FCI wave function covers therefore all electron correlation effects.

However, FCI calculations are only feasible for small molecules, since the number of excited determinants for m one-electron basis functions and N electrons is given by $\binom{2m}{N}$ [27]. For this reason, the FCI wave function is approximated to be able to perform calculations also for larger molecules. As a first step, the expansion of the FCI wave function is rewritten in a systematic way by grouping the CSFs according to the degree of orbital substitution (called excitation):

$$\Psi_m^{\mathrm{FCI}} = c_{m,0}\Phi_0 + \sum_{(a_j,i_j)} c_m, (a_j,i_j)\Phi_{a_j}^{i_j} + \sum_{(a_j,i_j),(a_k,i_k)} c_m,(a_j,i_j),(a_k,i_k)\Phi_{a_j,a_k}^{i_j,i_k} + \cdots, \tag{12}$$

with $\left\{\Phi_{a_j}^{i_j}\right\}$ denoting all singly excited determinants which can be obtained from the ground-state determinant, $\left\{\Phi_{a_j,a_k}^{i_j,i_k}\right\}$ all doubly excited determinants, and so forth. Obviously, the simplest approximation is obtained by truncating the expansion after a certain class of terms (truncated CI). Taking all singly excited determinants into account is referred to as CIS (CI-Singles), whereas incorporating also the doubly excited determinants yields CISD (CI-Singles-Doubles), and so forth. A major drawback is the violation of size consistency in the truncated CI approach, i.e., the energy of two identical molecules at infinite separation is not equal to two times the energy of a single molecule.

Different variants of truncated CI approaches exist. In truncated CI, the excited determinants are all obtained from the ground-state determinant. If also from other reference determinants excited determinants are produced, we arrive at the multi-reference CI (MRCI) approach. Sometimes, it is also convenient to define a restricted orbital space (active space), incorporate all excited determinants within this space, and reoptimize the orbital basis which is known as complete-active-space self-consistent-field approach (CASSCF).

There exists an approximation to the FCI wave function which overcomes the size-consistency problem of the truncated CI approach. This is the coupled-cluster approach. For the coupled-cluster approximation, we first define an excitation operator:

$$T = \sum_{k=1}^{\infty} T_k, \tag{13}$$

which contains all possible excitations from the ground-state determinant such that the operator T_k, when acting on the ground-state determinant, produces a linear combination of all possible k-fold excited determinants,

$$T_k\Theta_0 = \sum_{(a_1,i_1)\ldots(a_k,i_k)} t_{a_1,\ldots a_k}^{i_1,\ldots,i_k}\Theta_{a_1,\ldots,a_k}^{i_1,\ldots,i_k}, \tag{14}$$

with the expansion coefficients $t^{i_1,\ldots,i_k}_{a_1,\ldots,a_k}$ called cluster amplitudes (they can be related to the CI coefficients). The coupled-cluster wave function is then obtained through an exponential ansatz:

$$\Psi_0^{\mathrm{CC}} = \exp(T)\Theta_0 \tag{15}$$

which can be Taylor expanded:

$$\begin{aligned}\exp(T) &= \sum_{n=0}^{\infty} \frac{1}{n!} \left[\sum_{k=1}^{\infty} T_k\right]^n \\ &= 1 + T_1 + T_2 + \cdots + \frac{1}{2!}T_1^2 + T_1T_2 + \frac{1}{3!}T_1^3 + \cdots .\end{aligned} \tag{16}$$

If the summations are not truncated, the coupled-cluster wave function is equal to the FCI wave function. Even if the excitation operator T is truncated after single excitations $T = T_1$ (CCS) or double excitations $T = T_1 + T_2$ (CCSD), the expansion of the exponential function contains higher excitations than a corresponding truncated CI through products of excitation operators like T_1T_2, the so-called disconnected clusters. This is also the reason why the coupled-cluster method is size consistent even in its truncated form. Practical applications, using a truncated excitation operator, require expressions that contain a finite number of terms. This can be achieved using a Baker–Campell–Hausdorff expansion leading to a series of nested commutators such that the electronic energy is calculated as:

$$E = \left\langle \Phi_0 \middle| H + [H,T] + \frac{1}{2!}[[H,T],T] + \frac{1}{3!}[[[H,T],T],T] + \frac{1}{4!}[[[[H,T],T],T],T] \middle| \Phi_0 \right\rangle . \tag{17}$$

Because the coupled-cluster approach is nonlinear in the amplitudes as can be deduced from (16), the corresponding equations for the calculations of the amplitudes are solved by projection.

3.2 The Hamiltonian Operator

The Hamiltonian for a system of M nuclei and N electrons, including all one-electron terms and two-particle interactions, is given by:

$$\begin{aligned}H = &\sum_{I=1}^{M} t_{\mathrm{n}}(I) + \sum_{i=1}^{N} t_{\mathrm{e}}(i) + \sum_{I=1}^{M}\sum_{J=I+1}^{M} v_{\mathrm{nn}}(I,J) \\ &+ \sum_{i=1}^{N}\sum_{j=i+1}^{N} v_{\mathrm{ee}}(i,j) + \sum_{I=1}^{M}\sum_{i=1}^{N} v_{\mathrm{ne}}(I,i)\end{aligned} \tag{18}$$

with $t_n(I)$ and $t_e(i)$ denoting the kinetic energy operators for nuclei I and electrons i. $v_{nn}(I, J)$ is the repulsion energy operator between nuclei I and J, $v_{ee}(i, j)$ the electron–electron repulsion operator of electrons i and j, and $v_{ne}(I, J)$ describes the attraction of electron i and nucleus I. The explicit form of the expressions for the operators is defined through the basic physics on which an approximation to the exact Hamiltonian relies. When one aims at an adequate description of all elements in the periodic table, including heavy metal atoms like actinides, the Dirac–Coulomb–Breit (DCB) Hamiltonian is the most suitable choice for chemistry. It is deduced from the Einsteinian relativity principle using classical electromagnetic fields [22]. There may exist extreme cases, where quantum electrodynamical (QED) corrections play a role [i.e., where a quantization of the electromagnetic field (photons) is necessary], but we consider these cases to be unimportant for general chemistry.

For the DCB Hamiltonian, we first introduce the one-electron Dirac Hamiltonian $h_D(i)$ for an electron in an external potential. It contains all one-electron operators and the Coulomb interaction between the single electron and the nuclei (in Gaussian units; used throughout):

$$h_D(i) = c\boldsymbol{\alpha}_i \cdot \left(\boldsymbol{p}_i + \frac{e}{c}\boldsymbol{A}\right) + (\beta_i - 1)m_e c^2 - e\phi - \sum_{I=1}^{M} \frac{Z_I e^2}{|\boldsymbol{r}_i - \boldsymbol{R}_I|}. \tag{19}$$

External electromagnetic fields, represented by the vector potential $\boldsymbol{A}$ and the scalar potential ϕ, are introduced via minimal coupling which ensures Lorentz covariance of the one-electron Dirac equation of motion. For an isolated atom, molecule, or crystal, we have $\boldsymbol{A} = 0$ and $\phi = 0$. In the more general case of a system containing M nuclei and N electrons, the DCB Hamiltonian includes the one-electron Dirac Hamiltonian as:

$$\begin{aligned} H_{DCB} = &\sum_{I=1}^{M} \frac{\boldsymbol{p}_I^2}{2m_I} + \sum_{i=1}^{N} \left[c\boldsymbol{\alpha}_i \cdot \left(\boldsymbol{p}_i + \frac{e}{c}\boldsymbol{A}\right) + (\boldsymbol{\beta}_i - 1)m_e c^2 - e\phi - \sum_{I=1}^{M} \frac{Z_I e^2}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} \right] \\ &+ \sum_{I=1}^{M} \sum_{J=I+1}^{M} \frac{Z_I Z_J e^2}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} g(i,j) \end{aligned} \tag{20}$$

with I and i being the particle indices, $\boldsymbol{R}_I$ and $\boldsymbol{r}_i$ the coordinates, $Z_I \cdot e$ and $-e$ the charges, $\boldsymbol{p}_I$ and $\boldsymbol{p}_i$ the momenta, and m_I and m_e the masses of the nuclei and the electrons, respectively. The speed of light is denoted as c. $\boldsymbol{\alpha}_i$ is a Dirac $\boldsymbol{\alpha}$ matrix for the i-th electron, where $\boldsymbol{\alpha} = (\alpha_x, \alpha_y, \alpha_z)$ is a three-dimensional vector of four-by-four matrices:

$$\alpha_x = \begin{pmatrix} 0 & 0 & & \\ 0 & 0 & & \sigma_x \\ & & 0 & 0 \\ & \sigma_x & 0 & 0 \end{pmatrix}, \quad \alpha_y = \begin{pmatrix} 0 & 0 & & \\ 0 & 0 & & \sigma_y \\ & & 0 & 0 \\ & \sigma_y & 0 & 0 \end{pmatrix}, \quad \alpha_z = \begin{pmatrix} 0 & 0 & & \\ 0 & 0 & & \sigma_z \\ & & 0 & 0 \\ & \sigma_z & 0 & 0 \end{pmatrix}, \tag{21}$$

which are built from the Pauli spin matrices $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \; \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{22}$$

and $\boldsymbol{\beta}_i = \mathrm{diag}(1,1,-1,-1)$ is a diagonal matrix. There are different levels of approximation for the electron–electron interaction. The Coulomb–Breit operator reads [22]:

$$g(i,j) = \frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \left[1 - \frac{\boldsymbol{\alpha}_i \cdot \boldsymbol{\alpha}_j}{2} - \frac{\left(\boldsymbol{\alpha}_i \cdot \left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)\right)\left(\boldsymbol{\alpha}_j \cdot \left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)\right)}{2\left|\boldsymbol{r}_i - \boldsymbol{r}_j\right|^2} \right]. \tag{23}$$

It is often convenient to employ to Gaunt operator [28, 29] to approximate the Breit operator. A rigorous approximation to the DCB Hamiltonian is obtained in the nonrelativistic limit by substituting the Coulomb–Breit operator by the instantaneous Coulomb interaction (which we already adopted for the electron–nucleus interaction). The resulting DC Hamiltonian neglects retardation effects of the electron–electron interaction which arise from the transmission of the interaction due to the finite value of the speed of light.

Since the $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ parameters in the one-electron Dirac Hamiltonian have a 4×4 structure, it immediately follows that the one-electron functions have four components. Owing to the 2×2 superstructure of the Hamiltonian [see (21)], they are grouped into two 2-spinors, which are for historical reasons denoted as large (L) and small (S) component:

$$\psi_i = \begin{pmatrix} \psi_i^1 \\ \psi_i^2 \\ \psi_i^3 \\ \psi_i^4 \end{pmatrix} = \begin{pmatrix} \psi_i^{\mathrm{L}} \\ \psi_i^{\mathrm{S}} \end{pmatrix}. \tag{24}$$

Since the Dirac Hamiltonian possesses a 2×2 superstructure, any other operator O can also be expressed as:

$$O = \begin{pmatrix} O^{\mathrm{LL}} & O^{\mathrm{LS}} \\ O^{\mathrm{SL}} & O^{\mathrm{SS}} \end{pmatrix}. \tag{25}$$

The Dirac equation for a single electron yields two sets of solutions, namely positive (electronic) and negative (positronic) energy eigenstates. The electronic eigenstates describe either freely moving electrons or electrons that are bound by an external potential, whereas the positronic states lead to conceptual and practical difficulties. Dirac interpreted all the positronic states to be occupied by electrons such that the excitation of an electron from this "sea of electrons" produces

a positively charged "hole" (Dirac hole theory). With this interpretation, Dirac predicted the existence of an antiparticle, which contains the same mass as the electron but carries the opposite charge. In practice, the positronic states are not occupied in a quantum chemical calculation. The positronic states lead to the consequence that the Hamiltonian is no longer bounded from below because any electron can in principle take an infinitely large negative energy, which may result in a variational collapse during the optimization of the wave function. Another related pathology of the Dirac formalism is known as the Brown–Ravenhall [30] disease, or continuum dissolution. In fully relativistic calculations, these pathologies are circumvented by technical tricks that can be formalized in terms of iteratively optimized projection operators.

However, the positronic solutions are not relevant to chemistry, and their calculation is therefore an unnecessary burden. We now discuss two possibilities to avoid the calculation of these positronic states by decoupling the large and the small components of the spinor. Eliminating two components from the spinor yields efficient two-component methods. Either they can be projected out, e.g., using the generalized Douglas–Kroll–Hess (DKH) unitary transformation technique [31–33] or they are eliminated in the so-called regular approximations [34]. Recently, efficient four-component formulations of the unitary block-diagonalization in one shot hold promises for future routine applications [35–40]. However, low-order regular approximations and DKH schemes are sufficiently accurate, efficient, and well embedded in quantum chemistry program packages tailored for routine quantum chemical calculations (like, e.g., MOLCAS [41] or ADF [42, 43]).

The DKH transformation technique is based on an idea by Douglas and Kroll [44] which was later rediscovered and turned into a practical method by Hess [31]. The DKH formalism has then further been developed by Reiher and Wolf [32, 33, 45–49]. The general unitary DKH transformation block-diagonalizes the Dirac Hamiltonian:

$$h_{\mathrm{bd}} = U h_{\mathrm{D}} U^{\dagger} = \begin{pmatrix} h_{+} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & h_{-} \end{pmatrix}, \tag{26}$$

resulting in two decoupled two-component matrix operators h_+ and h_- for the electronic and the positronic eigenstates, respectively. The unitary transformation is constructed as a product of infinitely many unitary transformations $U = U_{\infty} \ldots U_3 U_2 U_1 U_0$ each Taylor expanded in terms of an antihermitian operator. In practical applications, the expansion of the unitary transformation is usually truncated.

Also the wave function $\tilde{\psi} = U\psi$ must be transformed in the same way. If expectation values of hermitian operators are calculated in the DKH picture, one must take care that the operators are transformed, too. The neglect of these transformations leads to the so-called picture change error (PCE) [50–52]. As a consequence, the electron density in the DKH picture is not obtained by simply

summing up weighted squares of the DKH orbitals. The PCE for any operator O is given by:

$$\mathrm{PCE(O)} = \left\langle \tilde{\Psi}^{\mathrm{L}} \middle| \tilde{O}^{\mathrm{LL}} \middle| \tilde{\Psi}^{\mathrm{L}} \right\rangle - \left\langle \tilde{\Psi}^{\mathrm{L}} \middle| O^{\mathrm{LL}} \middle| \tilde{\Psi}^{\mathrm{L}} \right\rangle, \tag{27}$$

where O^{LL} denotes the upper left block of O [see (25)], while $\tilde{O}^{\mathrm{LL}}$ denotes the upper left block of the *transformed operator* $\tilde{O}$. The notation reads then as DKH(n, m), with n being the order of the DKH transformation of the wave function and m the order of the DKH transformation of the property operator. If no property operator is used, the notation is simply given by DKHn, with n being the order of the transformation of the wave function.

A second way to reduce the four-component Dirac Hamiltonian to a two-component Hamiltonian is given by the regular approximation approach, which was introduced in 1986 by Heully et al. [53], Durand [54], and Chang et al. [55], and rediscovered by van Lenthe, van Leeuwen, Baerends, and Snijders [34, 56–59]. It relates the small component of the four-component wave function via the energy-dependent X-operator to the large component. The starting point for the regular approximation is the Dirac equation in split notation after applying an energy shift of $m_{\mathrm{e}}c^2$:

$$c\boldsymbol{\sigma}\cdot\boldsymbol{p}\psi_i^{\mathrm{S}} + V\psi_i^{\mathrm{L}} = \underbrace{\left(i\hbar\frac{\partial}{\partial t} - m_{\mathrm{e}}c^2\right)}_{\to\varepsilon_i}\psi_i^{\mathrm{L}}, \tag{28}$$

$$c\boldsymbol{\sigma}\cdot\boldsymbol{p}\psi_i^{\mathrm{L}} - 2m_{\mathrm{e}}c^2\psi_i^{\mathrm{S}} + V\psi_i^{\mathrm{S}} = \underbrace{\left(i\hbar\frac{\partial}{\partial t} - m_{\mathrm{e}}c^2\right)}_{\to\varepsilon_i}\psi_i^{\mathrm{S}}. \tag{29}$$

The energy-dependent $X(\varepsilon_i)$ operator is obtained from the lower part of the Dirac equation as:

$$\begin{aligned}\psi_i^{\mathrm{S}} = \underbrace{\frac{c\boldsymbol{\sigma}\cdot\boldsymbol{p}}{\varepsilon_i - V + 2m_{\mathrm{e}}c^2}}_{X(\varepsilon_i)}\psi_i^{\mathrm{L}} &= \frac{c\boldsymbol{\sigma}\cdot\boldsymbol{p}}{2m_{\mathrm{e}}c^2 - V}\left[1 - \frac{\varepsilon_i}{V - 2m_{\mathrm{e}}c^2}\right]^{-1}\psi_i^{\mathrm{L}} \\ &= \frac{c\boldsymbol{\sigma}\cdot\boldsymbol{p}}{2m_{\mathrm{e}}c^2 - V}\sum_{k=0}^{\infty}\left(\frac{\varepsilon_i}{V - 2m_{\mathrm{e}}c^2}\right)^k\psi_i^{\mathrm{L}}\end{aligned} \tag{30}$$

and can then in the next step be expanded in a geometric series with the expansion parameter $\varepsilon_i/(V - 2m_ec^2)$. The expanded form of the X-operator is then inserted in the upper part of the Dirac equation and yields:

$$\left(V + \frac{(\boldsymbol{\sigma}\cdot\boldsymbol{p})c^2}{2m_{\mathrm{e}}c^2 - V}\left[\sum_{k=0}^{\infty}\left(\frac{\varepsilon_i}{V - 2m_{\mathrm{e}}c^2}\right)^k\right](\boldsymbol{\sigma}\cdot\boldsymbol{p})\right)\psi_i^{\mathrm{L}}(\boldsymbol{r}) = \varepsilon_i\psi_i^{\mathrm{L}}(\boldsymbol{r}). \tag{31}$$

Truncation of the X-operator after the zeroth-order term $X_0 = c\boldsymbol{\sigma}\cdot\boldsymbol{p}/(2m_\mathrm{e}c^2 - V)$ leads then to the two-component ZORA Hamiltonian:

$$h_\mathrm{ZORA} = \boldsymbol{\sigma}\cdot\boldsymbol{p}\frac{c^2}{2m_\mathrm{e}c^2 - V}\boldsymbol{\sigma}\cdot\boldsymbol{p} + V. \tag{32}$$

The order of the regular approximation is thus determined by the order of the expansion of the energy-dependent X-operator in terms of $\varepsilon_i/(V - 2m_\mathrm{e}c^2)$. The derivation of the ZORA Hamiltonian is related to a Foldy–Wouthuysen transformation [60, 61]:

$$U_0 = \begin{pmatrix} \dfrac{1}{\sqrt{1 + X_0^\dagger X_0}} & \dfrac{1}{\sqrt{1 + X_0^\dagger X_0}}X_0^\dagger \\ -\dfrac{1}{\sqrt{1 + X_0 X_0^\dagger}}X_0 & \dfrac{1}{\sqrt{1 + X_0 X_0^\dagger}} \end{pmatrix}, \tag{33}$$

yielding

$$U_0 h_\mathrm{D} U_0^{-1} = \frac{1}{\sqrt{1 + X_0^\dagger X_0}} h_\mathrm{ZORA} \frac{1}{\sqrt{1 + X_0^\dagger X_0}}. \tag{34}$$

The ZORA Hamiltonian is then obtained by truncating the expansion of the reciprocal square root $1/\sqrt{1 + X_0^\dagger X_0} = 1 - 1/2X_0^\dagger X_0 + \cdots$ after the zeroth order term which is simply unity. Since only the four-component wave function is normalized and not the large component, it has to be normalized such that $\psi_\mathrm{ZORA} = \sqrt{1 + X_0^\dagger X_0}\psi^\mathrm{L}$ describes the normalized ZORA wave function.

The DKH and the ZORA Hamiltonian reduce the number of components in the wave function from four to two. So-called scalar-relativistic one-component methods are obtained by neglecting all spin-dependent terms in the DKH and the ZORA Hamiltonian. A further approximation, which also reduces the dimension to one, is given by the nonrelativistic limit of the Dirac equation, where the speed of light approaches infinity. The result is the one-component many-electron Schrödinger equation with the corresponding Hamiltonian:

$$\begin{aligned} H_\mathrm{non-rel.} = &\sum_{I=1}^{M}\frac{\boldsymbol{p}_I^2}{2m_I} + \sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m_\mathrm{e}} + \sum_{I=1}^{M}\sum_{J=I+1}^{M}\frac{Z_I Z_J e^2}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} \\ &+ \sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \sum_{i=1}^{N}\sum_{I=1}^{M}\frac{Z_I e^2}{|\boldsymbol{r}_i - \boldsymbol{R}_I|}, \end{aligned} \tag{35}$$

being the standard reference for nonrelativistic calculations.

3.3 Electron Density Distribution for Different Choices of Hamiltonian and Wave Function

The next step in the calculation of the electron density and the current density is to choose and insert a Hamiltonian operator and an ansatz for the wave function into the continuity equation (5), which contains the commutator of the Hamiltonian and the density operator. First, we discuss the case in which the many-electron wave function is approximated by a single Slater determinant. The electron density $\rho(\boldsymbol{r})$ is then calculated by inserting the definition of the Slater determinant given by (8). By applying the Slater–Condon rules [62, 63] for the calculation of matrix elements, the integrals containing the Slater determinants collapse to a sum of one-electron integrals:

$$\begin{aligned}\rho(\boldsymbol{r}) &= \langle \Theta(\{\boldsymbol{r}_i\})|\rho_{\boldsymbol{r}}|\Theta(\{\boldsymbol{r}_i\})\rangle = \left\langle \Theta(\{\boldsymbol{r}_i\}) \middle| \sum_{i=1}^{N} \delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}_i) \middle| \Theta(\{\boldsymbol{r}_i\}) \right\rangle \\ &= \sum_{i=1}^{N} \left\langle \psi_i(\boldsymbol{s}) \middle| \delta^{(3)}(\boldsymbol{r}-\boldsymbol{s}) \middle| \psi_i(\boldsymbol{s}) \right\rangle, \end{aligned} \tag{36}$$

for which we choose the arbitrary integration variable $\boldsymbol{s}$. The sum of these one-electron integrals yields the trace of the density matrix, whose diagonal elements define orbital densities:

$$\begin{aligned}\rho(\boldsymbol{r}) &= \sum_{i=1}^{N} \left\langle \psi_i(\boldsymbol{s}) \middle| \delta^{(3)}(\boldsymbol{r}-\boldsymbol{s}) \middle| \psi_i(\boldsymbol{s}) \right\rangle \\ &= \sum_{i=1}^{N} \psi_i^{\dagger}(\boldsymbol{r})\psi_i(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2, \end{aligned} \tag{37}$$

where $\psi_i^{\dagger}$ denotes the transposed and complex conjugate of the four-component orbital ψ_i (if ψ_i denotes a one-component orbital, then the dagger † is replaced by a star * for complex conjugation as transposing is inapplicable). Proceeding now with the insertion of the DCB Hamiltonian in the right-hand side of the continuity equation in (5), all multiplicative operators cancel in the commutator so that it reads:

$$\begin{aligned}\frac{\partial \langle \Psi|\rho_{\boldsymbol{r}}|\Psi\rangle}{\partial t} &= \frac{i}{\hbar}\left\langle \Psi \middle| \sum_{i=1}^{N} \left[c\boldsymbol{\alpha}_i \cdot \boldsymbol{p}_i, \delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}_i) \right] \middle| \Psi \right\rangle \\ &= -\nabla \cdot cN \left\langle \Psi \middle| \boldsymbol{\alpha}_1 \delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}_1) \middle| \Psi \right\rangle. \end{aligned} \tag{38}$$

From this equation, we can define the current density $\boldsymbol{j}_{\mathrm{DCB}} \stackrel{!}{=} cN\langle \Psi|\boldsymbol{\alpha}_1\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}_1)|\Psi\rangle$ for any type of four-component wave function. Hence, in the case of

a single Slater determinant (SD), the electron density and the current density are then given as:

$$\rho_{\mathrm{DCB}}^{\mathrm{SD}}(\boldsymbol{r}) = \sum_{i=1}^{N} \psi_i^{\dagger}(\boldsymbol{r}) \cdot \psi_i(\boldsymbol{r}) = \sum_{i=1}^{N} \left|\psi_i^{L}(\boldsymbol{r})\right|^2 + \left|\psi_i^{S}(\boldsymbol{r})\right|^2, \tag{39}$$

$$\boldsymbol{j}_{\mathrm{DCB}}^{\mathrm{SD}}(\boldsymbol{r}) = c \sum_{i=1}^{N} \psi_i^{\dagger}(\boldsymbol{r}) \cdot \boldsymbol{\alpha} \cdot \psi_i(\boldsymbol{r}). \tag{40}$$

In contrast to the DCB Hamiltonian, the electron density in the DKH framework is not obtained as the sum of squared DKH orbitals. The latter deviates from the electron density, especially in the region around the nucleus. At larger distances from the nucleus, these errors decrease [64, 65]. It is possible to relate the square of the Dirac 4-spinors to the square of the two-component DKH spinor by the introduction of a position-dependent error $\Delta\rho_i(\boldsymbol{r})$:

$$\left|\psi_{i,4\mathrm{comp}}(\boldsymbol{r})\right|^2 = \left|\psi_{i,\mathrm{DKH}}(\boldsymbol{r})\right|^2 + \Delta\rho_i(\boldsymbol{r}). \tag{41}$$

Considering the evaluation of expectation values over operators with a transformed wave function, one must not forget to transform the operator, too. The position-dependent error $\Delta\rho(\boldsymbol{r})$ in the calculation of the electron density vanishes when the transformed (picture-change corrected) density operator $\tilde{\rho}_{\boldsymbol{r}}$ is used [66]. The electron density obtained in this way is then equal to the Dirac electron density. The continuity equation is obtained in the same way as for the Dirac–Coulomb–Breit Hamiltonian, but with the transformed density operator and Hamiltonian applied. Then, the continuity equation reads:

$$\frac{\partial}{\partial t}\left\langle \boldsymbol{U}\Psi \middle| \boldsymbol{U}\rho_{\boldsymbol{r}}\boldsymbol{U}^{\dagger} \middle| \boldsymbol{U}\Psi \right\rangle = -\nabla \cdot cN\left\langle \boldsymbol{U}\Psi \middle| \boldsymbol{U}^{(1)}\boldsymbol{\alpha}_1\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}_1)\boldsymbol{U}^{(1)\dagger} \middle| \boldsymbol{U}\Psi \right\rangle. \tag{42}$$

The electron density and the current density are therefore obtained as:

$$\rho_{\mathrm{DKH}}^{\mathrm{SD}}(\boldsymbol{r}) = \sum_{i=1}^{N} \left\langle \tilde{\Psi} \middle| \boldsymbol{U}\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}_i)\boldsymbol{U}^{\dagger} \middle| \tilde{\Psi} \right\rangle, \tag{43}$$

$$\boldsymbol{j}_{\mathrm{DKH}}^{\mathrm{SD}}(\boldsymbol{r}) = c \sum_{i=1}^{N} \left\langle \tilde{\Psi} \middle| \boldsymbol{U}^{(i)}\boldsymbol{\alpha}_i\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}_i)\boldsymbol{U}^{(i)\dagger} \middle| \tilde{\Psi} \right\rangle. \tag{44}$$

In the case of more than one electron, the unitary transformation is given as the direct product of N unitary transformations $\boldsymbol{U} = \otimes_{i=1}^{N} U^{(i)}$.

Similar to DKH, the calculation of the electron density causes problems in the ZORA approach [50]. The electron density and the current density obtained from

$$\rho^{\mathrm{SD}}_{\mathrm{ZORA}}(\boldsymbol{r}) = \sum_{i=0}^{N} \psi^{\dagger}_{i,\mathrm{ZORA}}(\boldsymbol{r}) \cdot \psi_{i,\mathrm{ZORA}}(\boldsymbol{r}) \tag{45}$$

$$j^{\mathrm{SD}}_{\mathrm{ZORA}}(\boldsymbol{r}) = c \sum_{i=0}^{N} \psi^{\dagger}_{i,\mathrm{ZORA}}(\boldsymbol{r}) \cdot \alpha \cdot \psi_{i,\mathrm{ZORA}}(\boldsymbol{r}) \tag{46}$$

are only approximations to the Dirac densities, since the elimination of the small component causes a picture change and introduces therefore an error. An improved ZORA electron density (ZORA-4 density) is obtained by a backtransformation of the small component and the introduction of a scaling factor. Following van Lenthe and Baerends [50], a small and a large component density are defined for each orbital:

$$\rho^{\mathrm{S}}_{i}(\boldsymbol{r}) = \frac{c^2 \left(\boldsymbol{\sigma} \cdot \boldsymbol{p}_i \psi_{i,\mathrm{ZORA}}(\boldsymbol{r})\right)^{\dagger} \left(\boldsymbol{\sigma} \cdot \boldsymbol{p}_i \psi_{i,\mathrm{ZORA}}(\boldsymbol{r})\right)}{\left(2c^2 - V\right)^2}, \tag{47}$$

$$\rho^{\mathrm{L}}_{i}(\boldsymbol{r}) = \psi^{\dagger}_{i,\mathrm{ZORA}}(\boldsymbol{r}) \cdot \psi_{i,\mathrm{ZORA}}(\boldsymbol{r}). \tag{48}$$

The ZORA-4 density, which is normalized to one, is then calculated as:

$$\rho^{\mathrm{SD}}_{\mathrm{ZORA-4}} = \sum_{i=0}^{N} \frac{\rho^{\mathrm{L}}_{i} + \rho^{\mathrm{S}}_{i}}{1 + \int \rho^{\mathrm{S}}_{i} \mathrm{d}^3 r}. \tag{49}$$

Inserting the nonrelativistic Schrödinger Hamiltonian into the continuity equation:

$$\frac{\partial \langle \Psi | \rho_{\boldsymbol{r}} | \Psi \rangle}{\mathrm{d}t} = \frac{\hbar}{2 m_{\mathrm{e}} i} \sum_{i=1}^{N} \left[\left\langle \nabla_i^2 \Psi \middle| \delta^{(3)}(\boldsymbol{r} - \boldsymbol{r}_i) \middle| \Psi \right\rangle - \left\langle \Psi \middle| \delta^{(3)}(\boldsymbol{r} - \boldsymbol{r}_i) \middle| \nabla_i^2 \Psi \right\rangle \right] \tag{50}$$

yields the nonrelativistic (NR) electron density and current density. In the case of a wave function approximated by a single Slater determinant, it is given by:

$$\rho^{\mathrm{SD}}_{\mathrm{NR}}(\boldsymbol{r}) = \sum_{i=1}^{N} \psi_i^*(\boldsymbol{r}) \psi_i(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2, \tag{51}$$

$$j^{\mathrm{SD}}_{\mathrm{NR}}(\boldsymbol{r}) = \frac{\hbar}{2 m_{\mathrm{e}} i} \sum_{i=1}^{N} \left[\psi_i^*(\boldsymbol{r}) \nabla \psi_i(\boldsymbol{r}) - (\nabla \psi_i(\boldsymbol{r}))^* \psi_i(\boldsymbol{r}) \right]. \tag{52}$$

3.4 Methodological Aspects: Density Functional Theory and Current-Density Functional Theory

In the introduction, it was mentioned that the ground state of a system is fully determined not only by its ground-state electronic wave function Ψ_0 but also by its ground-state electron density $\rho_0(\boldsymbol{r})$. Since the wave function depends on $3N$ spatial coordinates (plus N additional spin coordinates), and since it is a fairly complicated object to handle, it would be preferable to have a theory which is solely based on the electron density, depending only on three spatial coordinates. For this reason, we give here a brief overview on DFT and its relativistic extension, namely current DFT, which are in the Kohn–Sham formulation both single-determinant methods. As discussed in the introduction, all properties of a system (especially the total energy) can be expressed as functionals of $\rho_0(\boldsymbol{r})$. The major drawback of DFT is that the analytical expressions for some of the energy contributions are not known, and thus some parts of the total energy must be approximated. Contemporary DFT is therefore not suited for highly accurate calculations, because the achievable accuracy strongly depends on the choice of the approximate density functionals. On the other hand, DFT is a simple, computationally not very demanding method which even includes electron correlation effects via an additional energy functional, and it allows one to perform calculations on large molecules.

The most widespread implementations of DFT are within the Kohn–Sham (KS) formalism [17], in which an artificial reference system of noninteracting electrons is introduced that yields exactly the same electron density as the interacting system. The energy contributions are partitioned in the following way (in Hartree atomic units):

$$\begin{aligned} E_{\mathrm{tot}}[\rho] &= T_{\mathrm{s}}[\rho] + V_{\mathrm{ext}}[\rho] + J[\rho] + E_{\mathrm{XC}}[\rho] + E_{\mathrm{NN}} \\ &= -\sum_{i=1}^{N/2} \left\langle \psi_i^{\mathrm{KS}} \middle| \nabla^2 \middle| \psi_i^{\mathrm{KS}} \right\rangle + \int \mathrm{d}^3 r\, \rho(\boldsymbol{r}) v_{\mathrm{ext}}(\boldsymbol{r}) \\ &\quad + \int\int \mathrm{d}^3 r\, \mathrm{d}^3 r' \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + E_{\mathrm{XC}}[\rho] + E_{\mathrm{NN}}. \end{aligned} \tag{53}$$

with $T_{\mathrm{s}}[\rho]$ denoting the kinetic energy of the noninteracting reference system, $V_{\mathrm{ext}}[\rho]$ the external potential energy which is caused by the nuclei, $J[\rho]$ the Coulomb interaction of the electrons, E_{NN} the nuclear repulsion energy, v_{ext} the external potential, and $E_{\mathrm{XC}}[\rho]$ being a sum of the nonclassical part of the electron–electron interaction and the difference between the kinetic energy of the noninteracting reference system and that of the interacting system. The exchange-correlation functional is the only unknown term in this expression, and hence its approximation determines the accuracy of the whole calculation. The (nonrelativistic) KS orbitals are calculated from the Kohn–Sham equations:

$$\left[-\frac{\hbar^2}{2m_\mathrm{e}}\nabla_i^2 + v_\mathrm{ext}(\boldsymbol{r}) + \int \mathrm{d}^3 r \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r}-\boldsymbol{r}'|} + \frac{\delta E_\mathrm{XC}[\rho]}{\delta\rho(\boldsymbol{r})}\right]\psi_i^\mathrm{KS}(\boldsymbol{r}) = \varepsilon_i \psi_i^\mathrm{KS}(\boldsymbol{r}), \tag{54}$$

which are obtained using the Lagrange method of undetermined multipliers to the variational procedure applied to the energy expression. The KS equations are one-electron equations that exhibit many similarities to the Hartree–Fock equations which are the basis of almost all wave function-based methods.

There exists a relativistic generalization of DFT called current DFT, which is based on fundamental relativistic concepts. The first requirement for this extension is a generalization of the Hohenberg–Kohn theorems which has been given by Rajagopal and Callaway [67]. The functionals then depend not only on the electron density $\rho(\boldsymbol{r})$ but also on the current density $\boldsymbol{j}(\boldsymbol{r})$. Both combine to the relativistic 4-current $j \equiv (c\rho, \boldsymbol{j})$. The total electronic energy in a four-component Kohn–Sham model then reads:

$$E_\mathrm{tot}[\rho,\boldsymbol{j}] = T_\mathrm{s}[\rho,\boldsymbol{j}] + V_\mathrm{ext}[\rho,\boldsymbol{j}] + J[\rho,\boldsymbol{r}] + E_\mathrm{XC}[\rho,\boldsymbol{j}]. \tag{55}$$

Following Rajagopal and Callaway [67], the most general form of the relativistic KS-DFT equations reads:

$$\left[c\boldsymbol{\alpha}\cdot\left(\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}_\mathrm{eff}(\boldsymbol{r})\right) + \boldsymbol{\beta} m_\mathrm{e} c^2 - e\phi_\mathrm{eff}(\boldsymbol{r})\right]\psi_i(\boldsymbol{r}) = \varepsilon_i \psi_i(\boldsymbol{r}), \tag{56}$$

where the Dirac Hamiltonian is clearly visible and supplemented by effective electromagnetic potentials defined as:

$$\begin{aligned}
\phi_\mathrm{eff}(\boldsymbol{r}) &= \phi_\mathrm{ext}(\boldsymbol{r}) - \frac{e}{c}\int \mathrm{d}^3 r' \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + c\frac{\delta E_\mathrm{XC}[\rho,\boldsymbol{j}]}{\delta\rho(\boldsymbol{r})}, \\
\boldsymbol{A}_\mathrm{eff}(\boldsymbol{r}) &= \boldsymbol{A}_\mathrm{ext}(\boldsymbol{r}) - \frac{e}{c}\int \mathrm{d}^3 r' \frac{\boldsymbol{j}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + c\frac{\delta E_\mathrm{XC}[\rho,\boldsymbol{j}]}{\delta\boldsymbol{j}(\boldsymbol{r})},
\end{aligned} \tag{57}$$

where the external potentials ϕ_ext and $\boldsymbol{A}_\mathrm{ext}$ represent the interaction between the electron and the nuclei in the case of isolated molecules.

4 Analysis of Approximate Electron Densities

Approximate wave functions and the specific choice of the Hamiltonian operator can cause errors in the calculation of the electron density, which have to be investigated systematically. The first part of this section focuses on the effect of approximate Hamiltonian operators, whereas the second part deals with the effects of the correlated motion of the electrons on the electron density.

4.1 Effect of the Approximate Hamiltonians

After the publication of the Dirac equation in 1928, Hartree [68] was the first to analyze the distribution of the charge and current in the Dirac formalism, but his publication did not contain any graphical representations of the electron density. The first graphical representation of the four-component Dirac electron density was then presented by White [69, 70], who investigated the angular distribution of the charge density for hydrogen-like atoms (highly positively charged one-electron ions). He presented normalized radial charge densities for different orbitals and made a few qualitative comments on the difference between the Schrödinger and the Dirac electron densities. After these pioneering studies, it took more than three decades until the first more detailed comparison between relativistic and nonrelativistic charge densities was presented by Burke and Grant [71] considering a hydrogen-like atom with a nuclear charge number $Z = 80$, i.e., Hg^{79+}. These authors compared the radial densities obtained with the Dirac–Coulomb Hamiltonian to nonrelativistic ones and drew general conclusions from it. The radial density $D(r)$ is given by:

$$D(r) = r^2 \int_0^{\pi} \int_0^{2\pi} \sin\,\vartheta \mathrm{d}\vartheta\,\mathrm{d}\varphi\, \Psi^{\dagger}(\boldsymbol{r})\Psi(\boldsymbol{r}) = P^2(r) + Q^2(r), \tag{58}$$

which is thus equal to the sum of the squares of the radial functions $Q(r)$ and $P(r)$ of the four-component wave function. The most general observation corresponds to a contraction of the relativistic density profiles toward the nucleus, which is most pronounced for the core-penetrating s- and $p_{1/2}$-shells. The degree of the contraction is affected by the absolute value of the relativistic azimuthal quantum number analog $|\kappa| = j + \frac{1}{2}$. For large $|\kappa|$ values, the relativistic density profiles resemble strongly the nonrelativistic ones, whereas the contraction is largest for $|\kappa| = 1$. A second observation concerns the nodes of the radial wave function, where the normalized radial electron density vanishes in the nonrelativistic case, but not in the relativistic one. The relativistic radial electron density is zero only at $r = 0$.

A number of publications appeared in the following years which investigated how relativistic effects affect the electron density [72–76], i.e., the differences between the Dirac and the Schrödinger picture. In 1993, van Lenthe, Baerends, and Snijders reported the regular approximations [34, 56, 57, 59] to the Dirac Hamiltonian and compared the r-weighted square root of the electron density for different orbitals of an uranium atom. For all of the outer shells, the relativistic contraction is fully recovered and $r\sqrt{\rho(\boldsymbol{r})}$ obtained from the ZORA Hamiltonian can hardly be distinguished from the Dirac results, except at the position of the minima, where it approaches zero in contrast to the Dirac result, which is always larger than zero. The only significant difference is observed for the innermost shell, where the ZORA Hamiltonian is able to recover a large part of the relativistic contraction, but not all of it. Furthermore, the maximum for the $1s_{1/2}$ orbitals is too large compared to the Dirac result. Another study by Autschbach and Schwarz [77]

compares the relativistic change of the radial density for hydrogen-like atoms with a nuclear charge number of $Z = 1$ and $Z = 100$. These authors compare the Dirac density to two different approximations, namely the Pauli (which is not discussed here as it is only of historical importance) and the ZORA Hamiltonians arriving at a similar conclusion as van Lenthe, Baerends, and Snijders.

DKH electron densities were presented by van Wüllen and Michauk [78] who emphasize in a brief discussion that the electron density obtained from the two-component wave functions (neglecting picture-change effects) is not equal to the Dirac density due to the picture-change error. As explained above, the DKH electron density, approximated by the sum of the squared DKH orbitals, is sufficiently accurate for the valence region of an atom, but the error increases, the smaller the distance from the nucleus is (this subject is further analyzed in Sect. 5). Picture-change-affected DKH density were also analyzed in [65], considering a hydrogen-like mercury atom with $Z = 80$.

Eickerling et al. [64] presented the first systematic investigation of the effects of an approximate two-component Hamiltonian and the scalar-relativistic DKH10 Hamiltonian on the electron density and its topology by comparing to the four-component Dirac Hamiltonian and to the nonrelativistic limit, namely the Schrödinger Hamiltonian. The study features a comparison of difference electron densities obtained from three relativistic and the nonrelativistic Hamiltonian for a homologous series of acetylene complexes M–C_2H_2 with M = Ni,Pd,Pt and in addition an analysis of the negative Laplacian at the bond critical points (BCPs), which are minima of the electron density on the bonding axis and maxima on the axis perpendicular to the bonding axis. The most significant difference between the four-component Dirac density and the nonrelativistic one, considering all BCPs, is observed for the M–C_2H_2 BCP in the case of M = Pt, where it amounts to 0.06 eÅ^{-3}. The study concludes that scalar-relativistic methods cover most relativistic effects, though there are still differences to the electron densities obtained from two-component methods. Concerning the negative Laplacians, the deviations can be larger which makes it a more sensitive measure than the electron density itself. Furthermore, the size of the relativistic effects is estimated to be of almost the same size as correlation effects in four-component DFT calculations.

In order to provide a closer look at the accuracy of approximate relativistic Hamiltonians, we discuss the results for the homologous acetylene complexes given by Eickerling et al. [64] in more detail. The difference electron densities $\rho_{\text{rel.}}(\boldsymbol{r}) - \rho_{\text{nonrel.}}(\boldsymbol{r})$ obtained from four-component Dirac, ZORA-SO (including spin–orbit effects), and scalar-relativistic DKH10 calculations are shown in Fig. 1a, d, g. Although the relativistic effects are expected to be most pronounced for the case of M = Pt, one can observe significant differences even for the nickel complex. The difference electron density map for the nickel complex contains different circular minima and maxima due to the changes in the radial extension of the atomic sub-shells. The innermost circular region of positive difference exhibits four maxima around the nickel atom, from which the one oriented in the direction of the acetylene ligand is more pronounced in the case of DKH10 than for the ZORA-SO or the four-component Dirac density. These maxima correspond to local charge

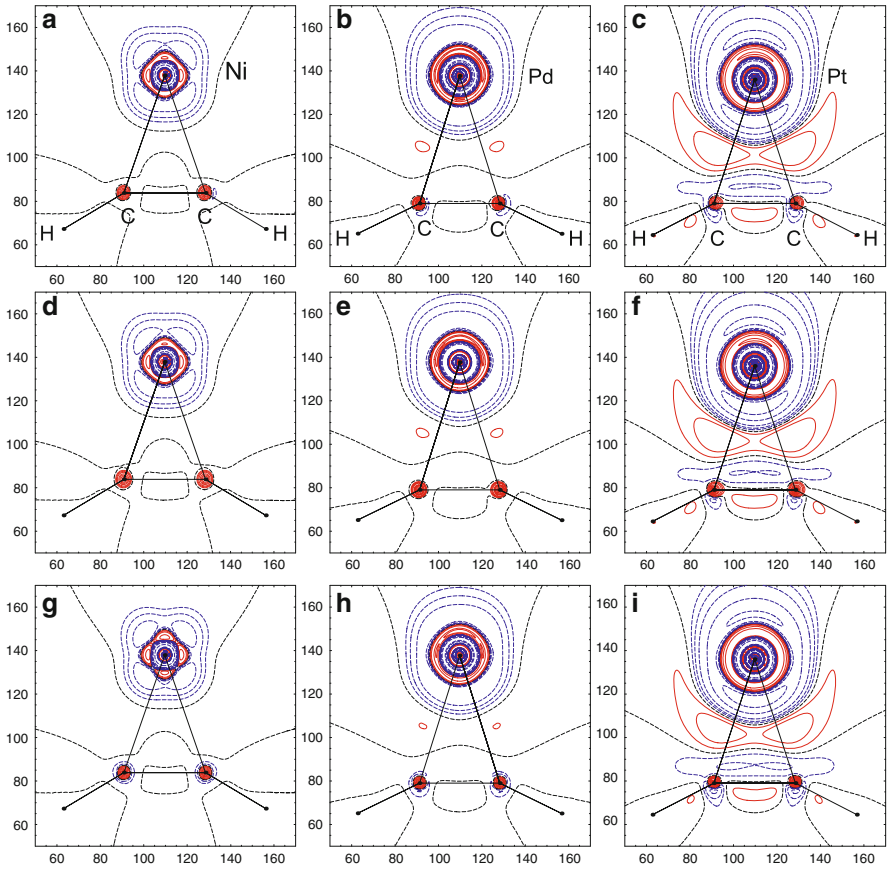


Fig. 1 Difference densities in the molecular plane, $\rho_{4comp}(\boldsymbol{r}) - \rho_{nonrel}(\boldsymbol{r})$ for (**a**) $Ni(C_2H_2)$, (**b**) $Pd(C_2H_2)$, (**c**) $Pt(C_2H_2)$; $\rho_{ZORA}(\boldsymbol{r}) - \rho_{nonrel}(\boldsymbol{r})$ for (**d**) $Ni(C_2H_2)$, (**e**) $Pd(C_2H_2)$, (**f**) $Pt(C_2H_2)$; and $\rho_{DKH10}(\boldsymbol{r}) - \rho_{nonrel}(\boldsymbol{r})$ for (**g**) $Ni(C_2H_2)$, (**h**) $Pd(C_2H_2)$, (**i**) $Pt(C_2H_2)$. Values of positive and negative difference densities are indicated by *solid* and *dashed lines*, respectively. Contour lines are drawn at $\pm 2, \pm 4, \pm 6, \pm 8 \times 10^n$ eÅ^{-3} with $n = 0, 1, 2$. Note that the axes labels denote grid points. (The figure is reprinted with permission from [64]. Copyright 2010 American Chemical Society)

concentrations of the valence region of the nickel atom. Moreover, the scalar-relativistic DKH10 Hamiltonian yields two more contour lines of negative difference than the Dirac or the ZORA-SO difference densities in the region around the carbon atoms of the acetylene ligand. From the results of the nickel complex, one obtains a first indication that the scalar-relativistic DKH10 Hamiltonian recovers less of the relativistic effects than the ZORA-SO Hamiltonian.

The difference density maps for the palladium complex are depicted in Fig. 1b, e, h. Here the situation is very similar to the nickel complex, but the relativistic effects are more pronounced. The four maxima in the valence region of the metal

atom are also present in this case, and two additional maxima are observed in the metal–ligand binding region that are not present in the nickel complex. The scalar-relativistic DKH10 results show deficiencies in the maximum which is oriented toward the acetylene ligand as for the case of the nickel complex, and also the two maxima in the bonding region of the Dirac difference density map are better recovered by the ZORA-SO result than by the DKH10 one.

The difference density for M = Pt is shown in Fig. 1c, f, i. As expected, the relativistic effects are the largest for this complex, which contains only one maximum in the positive difference region around the Pt atom, on the opposite site of the acetylene ligand. This maximum is only present in the case of the four-component Dirac and the ZORA-SO result, but not in the scalar-relativistic DKH10 one, which points to the importance of spin–orbit effects on the electron density. When comparing to the other two complexes, the maximum is almost one order of magnitude larger.

4.2 *Significance of Electron Correlation: $Fe(NO)^{2+}$ as an Example*

In Sect. 3.1, it was mentioned that an independent-particle model used for the electronic wave function (Hartree–Fock theory) does not consider effects in the wave function which arise due to the correlated motion of the electrons. Early studies analyzing correlation effects on the electron density were presented by Bader and Chandra [79] for the H_2 molecule, which was also subject of a later study by Baerends et al. [80]. Bader and Chandra compared electron densities obtained from extended Hartree–Fock and Hartree–Fock (HF) calculations to understand how correlation effects affect the electron density. Following their paper, the electron density is in the case of HF overestimated in the central bonding region, whereas it is underestimated in the region around the nuclei. The authors also presented difference density plots for the Li_2 molecule arriving at the conclusion that in this case, the correlation effects on the charge density are negligible, because they are of the same magnitude as the accuracy of the density distribution itself. Smith [81] extended the study of Bader and Chandra, incorporating H_3^+ in a comparison of CI and HF electron densities, regarding also the difference between the atomic densities obtained from these calculations.

During the 1980s, various studies [82–85] presented correlation densities of systems larger than H_2. Stephens et al. [82] analyzed the influence of electron correlation on the partitioning of the electron density into atomic contributions (a decomposition to be discussed in Sect. 6.2) using BeO and CO as model systems. The study states that the so-called zero-flux surfaces (compare Sect. 6.2) are not very much affected by electron correlation. Gatti et al. [85] started then a systematical study investigating correlation effects on the charge density as well as on the Laplacian and on atomic properties of many three-atomic molecules. These authors

compared the electron density obtained from CI calculations to HF calculations, drawing similar conclusions as Bader and Chandra [79]. The electron density in the bonding region is decreased by the consideration of correlation effects.

In 1991, Kraka, Gauss, and Cremer [86] undertook the systematic investigation of correlation effects in Moller–Plesset perturbation theory of n-th order (MPn). They took into account calculations up to fifth order and compared also MP2 to HF for the CO molecule. The correlation corrections to the electron density oscillate up to fifth order. The effects are smaller at the equilibrium geometry and grow with interatomic distance. Cremer and He [87, 88] then published two further studies that consider electron correlation as covered by DFT in comparison to wave function-based methods. Within the local density approximation (LDA), electron density is enhanced in the bonding region and around the nuclei. Compared to gradient-corrected density functional calculations, the effects of the LDA functional are partially reduced, because electron density is shifted back from the bonding region and the region around the nuclei into the valence region of the molecule.

The study of Eickerling et al. [64] also contains a part that considers correlation effects on the topology of the electron density. The authors use the same model systems as for the study of relativistic effects and present a comparison with results obtained from Dirac–Hartree–Fock (DHF) and DFT calculations. Incorporation of correlation effects lowers the values of the density at the C–C and the C–H BCPs for all model systems, which is in good agreement with the studies of Bader and Chandra [79] and of Gatti et al. [85]. For the case of the M–C BCPs, there is no clear trend visible. Considering M = Pt and M = Ni, the electron density is also lowered at this BCP, whereas due to an error cancellation the values for M = Pd are almost equal for DHF and DFT.

Two very recent papers by Jankowski et al. [89, 90] investigate dynamical correlation effects on the electron density for DFT calculations considering the noble gas atoms neon and argon. The authors state that even though dynamical correlation effects on the electron density are weak, the shape of the curves is very sensitive to the changes in the electron density. Dynamical correlation effects are not well represented by density functionals which contain either the VWN5 [91] or the LYP [92] correlation functional. Better results are obtained when orbital-dependent OEP2-f [93] correlation functionals are used.

We shall here present new results for correlation densities [$\rho_{\mathrm{CASSCF}}(\boldsymbol{r}) - \rho_{\mathrm{HF}}(\boldsymbol{r})$, and $\rho_{\mathrm{DFT}}(\boldsymbol{r}) - \rho_{\mathrm{HF}}(\boldsymbol{r})$] obtained from CASSCF and DFT calculations for *transition metal complexes*. We choose a linear and a bent $Fe(NO)^{2+}$ structure as model systems. The result from a CAS(13,13) calculation shall serve as reference density, in which static correlation is included, but the CAS can be considered sufficiently large to cover also a substantial amount of dynamic correlation as is evident from a detailed study of the spin density of $[Fe(NO)]^{2+}$ [94] (for the consideration of additional dynamic correlation effects CASPT2 calculations would be required). Considering the Hartree–Fock orbital energies, the largest CAS feasible was chosen, including the 13 orbitals depicted in Fig. 2 for the linear complex and in Fig. 3 for the bent one. The aim of this investigation is to finally compare DFT results to

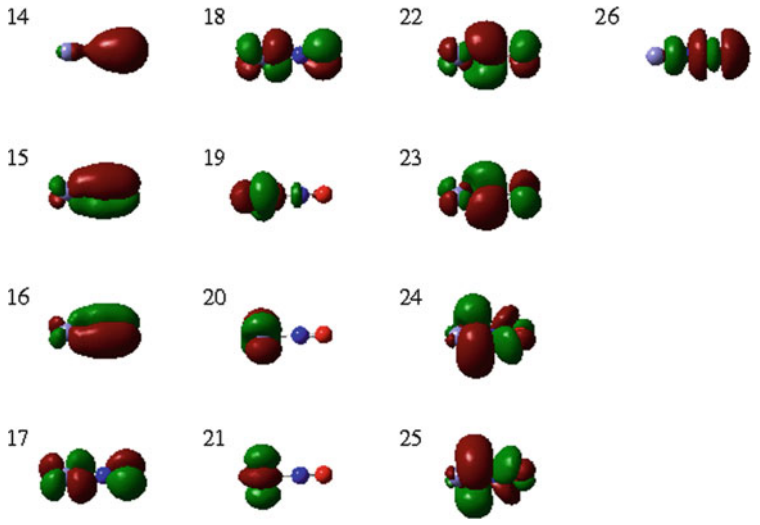


Fig. 2 DKH2-HF molecular orbitals 14–26 of the linear $Fe(NO)^{2+}$ complex, which were used in the CAS(13,13) calculation

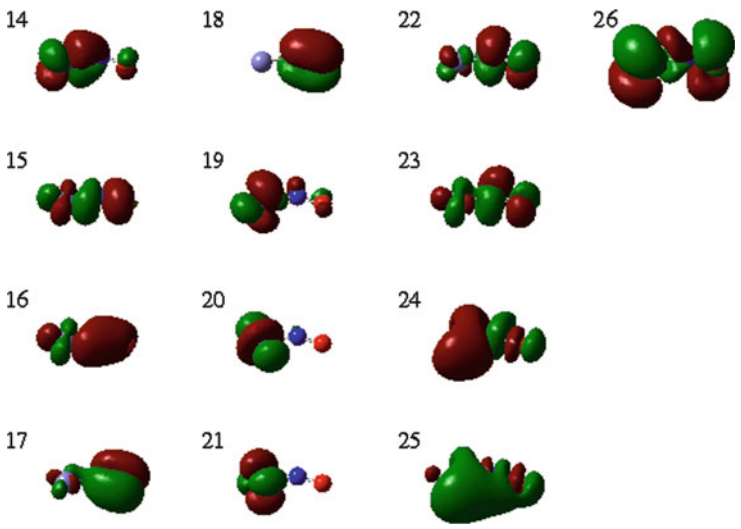


Fig. 3 DKH2-HF molecular orbitals 14–26 of the bent $Fe(NO)^{2+}$ complex, which were used in the CAS(13,13) calculation

the CASSCF reference to analyze to which extent correlation effects are covered by (semi-empirical) DFT.

The correlation density $\rho_{\mathrm{CASSCF}}(\boldsymbol{r}) - \rho_{\mathrm{HF}}(\boldsymbol{r})$ in the plane of the bonding axis for the linear nitrosyl complex is shown in Fig. 4a. It takes values between -0.10 and

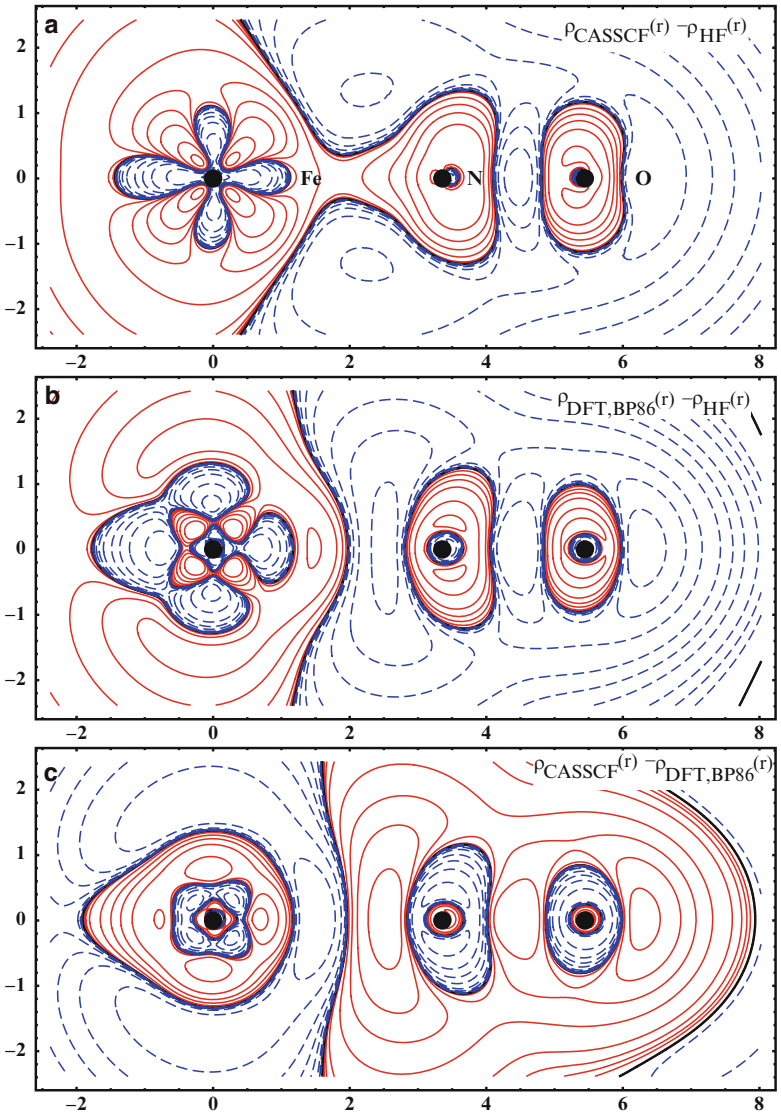


Fig. 4 Difference densities along the bonding axis of a linear $Fe(NO)^{2+}$ complex, for DKH(2,0) calculations. (**a**) $\rho_{CASSCF}(\boldsymbol{r}) - \rho_{HF}(\boldsymbol{r})$, (**b**) $\rho_{DFT,BP86}(\boldsymbol{r}) - \rho_{HF}(\boldsymbol{r})$, and (**c**) $\rho_{CASSCF}(\boldsymbol{r}) - \rho_{DFT,BP86}(\boldsymbol{r})$. Values of positive difference correspond to *red solid lines*, whereas values of negative difference are indicated by *blue dashed lines* and the *solid black line* denotes the region of zero-difference. Contour lines are drawn at $\pm 2, \pm 4, \pm 8 \times 10^n$ eÅ^{-3} with $n = -2, -3, -4, -5$. The contour values start at $\pm 2 \times 10^{-5}$ eÅ^{-3} and are then always doubled until $\pm 8 \times 10^{-2}$ eÅ^{-3}

0.05 e/Bohr^3 which already shows that the correlation effects on the electron density are not very large as noted in the previous studies [89, 90]. Around the iron atom, the correlation density exhibits four minima on the bonding axis and the

axis perpendicular to it. These minima are surrounded by four maxima, such that one can conclude that electron density is shifted from the bonding axis to regions around the nuclei. The correlation density of the nitrogen and the oxygen atoms

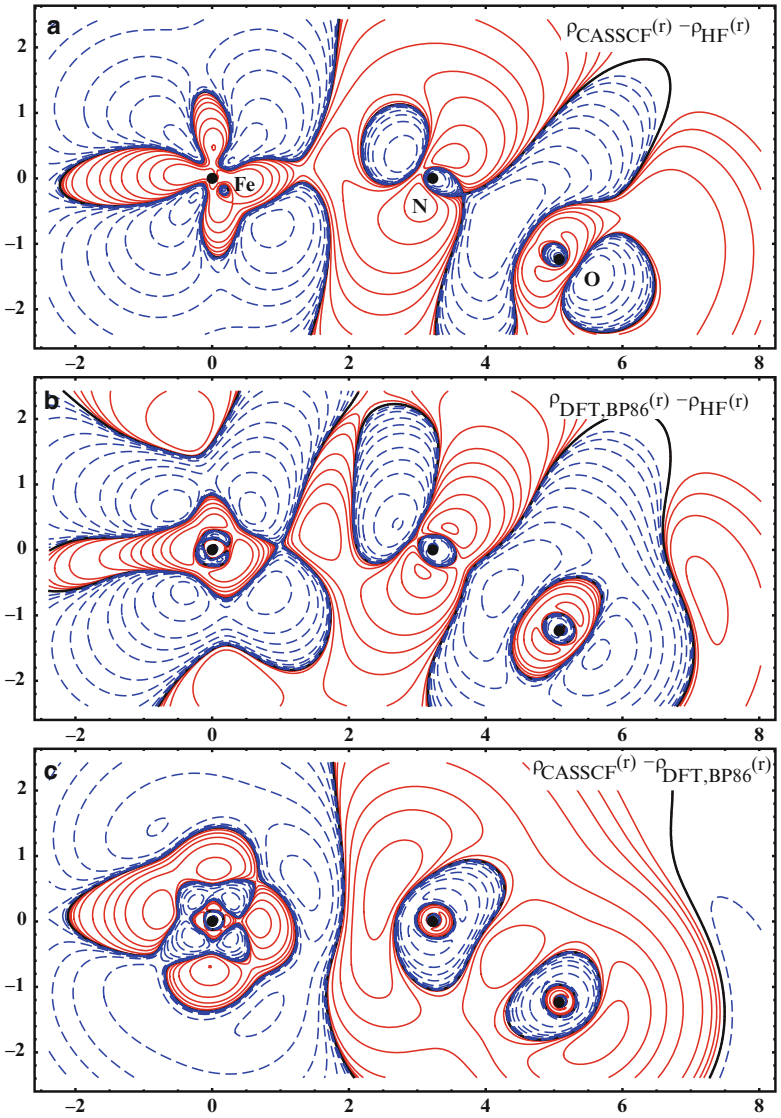


Fig. 5 Difference densities along the bonding axis of a bent $Fe(NO)^{2+}$, for DKH(2,0) calculations. (**a**) $\rho_{CASSCF}(\boldsymbol{r}) - \rho_{HF}(\boldsymbol{r})$, (**b**) $\rho_{DFT,BP86}(\boldsymbol{r}) - \rho_{HF}(\boldsymbol{r})$, and (**c**) $\rho_{CASSCF}(\boldsymbol{r}) - \rho_{DFT,BP86}(\boldsymbol{r})$. Values of positive difference correspond to *red solid lines*, whereas values of negative difference are indicated by *blue dashed lines* and the *solid black line* denotes the region of zero-difference. Contour lines are drawn at ± 2, ± 4, $\pm 8 \times 10^n$ eÅ^{-3} with $n = -2, -3, -4, -5$. The contour values start at $\pm 2 \times 10^{-5}$ eÅ^{-3} and are then always doubled until $\pm 8 \times 10^{-2}$ eÅ^{-3}

shows in contrast to the iron atom a different behavior. Here the difference density exhibits minima on the axis perpendicular to the bonding axis, but for larger distances also maxima can be observed. There is one minimum on the bonding axis between the nitrogen and the oxygen atom. The difference density $\rho_{\mathrm{DFT}}(\boldsymbol{r}) - \rho_{\mathrm{HF}}(\boldsymbol{r})$ and the difference between both correlation densities are shown in Fig. 4b, c. In general, DFT recovers most of the regions of positive and negative differences from the CASSCF correlation density. At the four maxima around the iron atom, DFT overestimates the correlation density, whereas the minima on the bonding axis are broader than in the case of the CASSCF reference. Summarizing our results, it can be stated that the major part of the differences in the correlation density is located directly around the atoms, but there are also four regions close to the iron nucleus where the correlation density vanishes totally. At larger distances, the correlation density tends to decrease quickly.

In the case of the bent structure, the correlation density $\rho_{\mathrm{CASSCF}}(\boldsymbol{r}) - \rho_{\mathrm{HF}}(\boldsymbol{r})$ along the bonding axis is shown in Fig. 5a. The absolute size of the correlation effects on the electron density is comparable to the one for the linear geometry. In the region around the iron atom, the situation is reversed, as the maxima are now located on the bonding axis and the axis perpendicular to it. Considering the nitrogen and the oxygen atoms, part of the electron density is shifted from the N–O bonding region to the nuclei as for the linear geometry. In addition, electron density is shifted from both ends of the nitrosyl ligand to the axis perpendicular to the N–O bonding axis. The difference density $\rho_{\mathrm{DFT}}(\boldsymbol{r}) - \rho_{\mathrm{HF}}(\boldsymbol{r})$ and the difference between both correlation densities are shown in Fig. 5b, c. As for the linear geometry, DFT recovers the correlation density reasonably well, except of some deficiencies at the iron and the oxygen atom. Therefore, the difference density distributions around the nuclei are almost identical when comparing CASSCF and DFT for both geometries. They exhibit maxima directly at the nuclei, surrounded by minima at intermediate distances, followed again by maxima which are located on the bonding axis and perpendicular to it.

5 The Most Difficult Case: Contact Densities

The picture-change-affected DKH electron density and the electron density obtained with the ZORA approach exhibit large deficiencies at the position of the nucleus. The electron density at this position is called contact density and plays an important role in the model theory of many spectroscopic techniques. An adequate theoretical description of the field shift in electronic transitions in high-resolution atomic electron spectra (first achieved by Ehrenfest [95, 96] and further developed by Rosenthal et al. [97] and Breit [98]) is for instance closely related to the difference of the contact densities of the atom for both electronic states that are involved in the transition. The isotopic field shift in the rotational spectra of a diatomic molecule is proportional to the first derivative of the contact density with respect to the equilibrium distance of the nuclei, whereas the isotopic field shift in

the vibrational energy depends on the second derivative [99, 100]. The contact density plays also an important role in Mössbauer spectroscopy, because it can be related to the chemical isomer shift [73, 74, 101], which was discovered in 1960 [102]. It depends on the differences of the contact density of the emitter source and the probe [103]. A more detailed description of the basic principles of Mössbauer spectroscopy and contact densities can be found in [104]. In order to be able to calculate these properties for atoms of the whole periodic table, one must ensure that the contact density can be calculated with sufficient accuracy.

The contact density is dominated by the core-penetrating *s*-orbitals that are strongly affected by relativistic effects. An adequate description requires therefore a fully relativistic treatment of the electrons. In addition, the choice of the nuclear charge distribution model has an effect on the accuracy of the calculated electron density. In the nonrelativistic picture, the nuclei are considered to be point charges, because it can be shown that the different nuclear charge distribution models usually yield negligible energy differences when compared to point-like nuclei – even for heavy nuclei [105]. The electron density features in this case a cusp at the location of the nucleus which is described by Kato's cusp condition [106]. In a fully relativistic description, the choice of the nuclear charge distribution model becomes important. For heavy atoms, the point-like description of the nuclei causes significant errors, whereas the differences between the available nuclear charge distribution models are small. For a comprehensive review on the available models, we recommend [107]. There are in general two possibilities to include these models into calculations. Either the calculation is performed for point-like nuclei and effects of finite nuclei are considered via perturbation theory, or they are directly incorporated in the calculation. The perturbation theory approach can be found in early studies [97, 98] on contact densities, whereas the direct consideration should be more consistent [108, 109]. For the major part of the practical applications, it is sufficient to use a simple model for the positive charge distribution in the atomic nucleus [105, 110]. Two simple models are the homogeneous charge distribution and the Gaussian charge distribution [108]. Also the three parameter Fermi model [111] has been used.

The first attempts to obtain fully relativistic electron density distributions considering finite-nucleus effects can be dated back to Rosenthal et al. [97] and Breit [98], who included the effects of a finite nuclear charge distribution via perturbation theory. A direct treatment of the finite nuclear charge distribution model using a three parameter Fermi model was then first presented by Fricke and Waber [108], who calculated Mössbauer isomer shifts.

In the case of solids, properties are calculated within solid-state density functional theory, where mostly the simple model of an uniformly charged sphere is used, e.g., as shown by Svane and Antoncik [112]. A recent study by Mastalerz et al. [109] investigated the basis set convergence for the calculation of contact densities at finite nuclei. A most recent fully relativistic investigation of correlation effects on the contact density with CCSD(T) was presented for mercury compounds in [113]. As a final remark to the calculation of the isomer shift, we refer to an alternative

way to calculate it within DFT by expressing it as a function of the derivative of the total energy with respect to the nuclear radius [101, 114].

In the context of approximate relativistic Hamiltonians, Mastalerz et al. [66] analyzed the DKH electron density at the position of the nucleus for different orders of the DKH transformation and investigate the effect of picture-change corrections and different nuclear charge distribution models. We select some results of this study considering the hydrogen-like Hg^{79+} highly charged ion for a more detailed discussion. Because the study [66] focused on the electron density at the position of the nucleus, the authors decided to analyze the $1/r^2$ weighted radial density [of (58)], which is depicted in Fig. 6 for the Hg^{79+} ion. The authors compared the weighted radial density obtained from fully numerical four-component and scalar-relativistic DKH(2,0), DKH(5,0), DKH(2,2), and DKH(9,9) calculations. The curves for point-charge nuclei feature in both cases [Dirac and DKH(n,0)] the weak singularity at the position of the nucleus, and the position-dependent error is clearly visible. The DKH(2,0) and DKH(5,0) curves cannot be distinguished from each other. Hence, already a low-order transformation of the orbitals is very accurate. In the case of a finite-nucleus model, the weighted radial density approaches finite values at the position of the nucleus in all cases, which can easily

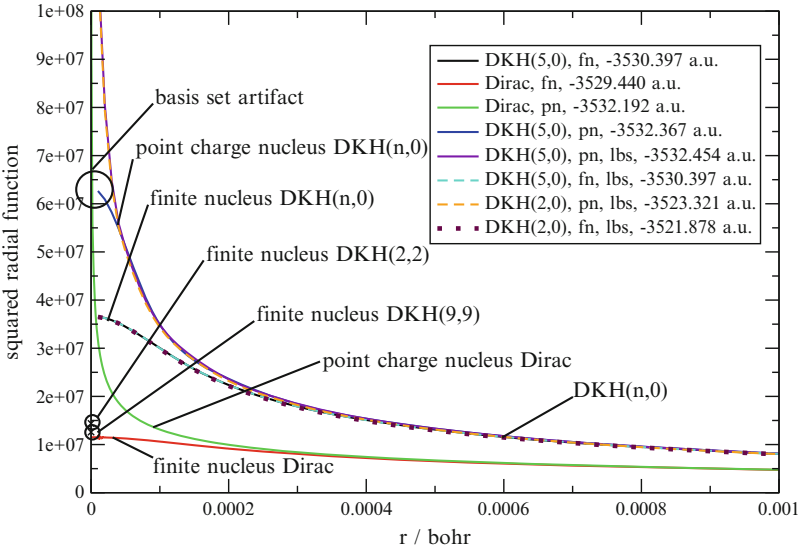


Fig. 6 $1/r^2$-weighted radial electron densities $\left[\rho_{1s_{1/2}}(\boldsymbol{r}) = P^2_{1s_{1/2}}(\boldsymbol{r})/\boldsymbol{r^2} + Q^2_{1s_{1/2}}(\boldsymbol{r})/\boldsymbol{r^2}(\text{Dirac})\right.$ and $\left.\rho_{1s}(\boldsymbol{r}) = P^2_{1s}(\boldsymbol{r})/\boldsymbol{r^2}(\text{DKH})(n,0)\right]$ of Hg^{+79} calculated with different nuclear models and basis sets. The DKH(n,0) curves are picture-change affected. The abbreviations "pn" and "fn" denote a point-charge nucleus and a finite Gaussian charge distribution model, respectively. "lbs" indicates that the calculation has been performed with a larger basis set that includes more steep functions. For comparison, the electronic energy in Hartree atomic units is given. Moreover, picture-change-corrected DKH(2,2) and DKH(9,9) data of the contact density is included. (The figure is reprinted with permission from [66]. Copyright 2010 Elsevier)

be reproduced by Gauss-type basis functions. The picture-change-affected DKH (n,0) curves exhibit large errors in the close vicinity of the nucleus and are located above the Dirac curve, whereas the picture-change-corrected DKH(n,n) contact densities match the Dirac contact densities [as would the picture-change-corrected DKH(n,n) and Dirac curves], even for the second-order DKH transformation of the property operator.

6 Electron Density in Conceptual Theories

6.1 Conceptual Density Functional Theory

Many principles of chemical reactivity such as electronegativity, chemical hardness/softness and frontier orbital theory had been proposed mostly as rather phenomenological descriptive concepts some decades ago. Many pioneering studies discussed for instance the Mulliken electronegativity [115–118], the hard and soft acid and base (HSAB) principle [119–123], and frontier molecular orbital concepts [124–129]. These concepts were then later unified and rigorously defined in the framework of conceptual DFT. Conceptual DFT was defined by Parr and coworkers [130], who recognized that the electronegativity can be defined as the negative of the derivative of the energy with respect to the number of electrons for a constant external potential. A detailed description of the history of conceptual DFT and various reactivity descriptors can be found in [131]. In principle, conceptual DFT relies on a Taylor series expansion of the total-energy density functional [132] with respect to perturbations in the external potential $v(\boldsymbol{r})$ and the number of electrons N:

$$\begin{aligned} &E[v(\boldsymbol{r})+\Delta v(\boldsymbol{r}),N+\Delta N] \\ &\quad = E[v(\boldsymbol{r}),N]+\Delta N\left(\frac{\partial E}{\partial N}\right)_{v(\mathbf{r})}+\int\left(\frac{\delta E}{\delta v(\boldsymbol{r})}\right)_N \Delta v(\boldsymbol{r})\mathrm{d}^3r+\frac{1}{2}(\Delta N)^2\left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\mathbf{r})} \\ &\quad +\frac{1}{2}\iint\left(\frac{\delta^2 E}{\delta v(\boldsymbol{r})\delta v(\boldsymbol{r}')}\right)_N(\Delta v(\boldsymbol{r}))^2\mathrm{d}^3r\mathrm{d}^3r'+\Delta N\int\left(\frac{\delta\partial E}{\delta v(\boldsymbol{r})\partial N}\right)\Delta v(\boldsymbol{r})\mathrm{d}^3r+\cdots. \end{aligned} \tag{59}$$

The external potential and the number of electrons are both functionals of the electron density. The derivatives of the energy with respect to either the external potential or the number of electrons or both of them can be identified with several chemical descriptors, shown in Table 1, that are related to chemical reactivity.

The mixed derivative which is first order in the external potential and in the number of electrons is called the Fukui function $f(\boldsymbol{r})$. It can be interpreted in two different ways. The definition as a derivative of the electron density implies that it represents the change in the electron density $\rho(\boldsymbol{r})$ at the point $\boldsymbol{r}$ when the number of particles N changes. However, the Fukui function characterizes also the response of the chemical

Table 1 Descriptors defined within conceptual DFT as energy derivatives

Symbol	Descriptor	Energy derivative
μ	Chemical potential	$\left(\frac{\partial E}{\partial N}\right)_{v(\mathbf{r})}$
χ	Electronegativity	$-\left(\frac{\partial E}{\partial N}\right)_{v(\mathbf{r})}$
$\rho(\boldsymbol{r})$	Electron density	$\left(\frac{\delta E}{\delta v(\boldsymbol{r})}\right)_{N}$
$\chi(\boldsymbol{r},\boldsymbol{r}')$	Linear response function	$\left(\frac{\delta^2 E}{\delta v(\boldsymbol{r})\delta v(\boldsymbol{r}')}\right)_{N} = \left(\frac{\delta \rho(\boldsymbol{r})}{\delta v(\boldsymbol{r})}\right)_{N}$
η	Chemical hardness	$\left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\mathbf{r})} = \left(\frac{\partial \mu}{\partial N}\right)_{v(\mathbf{r})}$
S	Chemical softness	$\frac{1}{\eta} = \left(\frac{\partial N}{\partial \mu}\right)_{v(\mathbf{r})}$
$f(\boldsymbol{r})$	Fukui function	$\left(\frac{\delta \partial E}{\delta v(\boldsymbol{r}) \partial N}\right) = \left(\frac{\partial \rho(\boldsymbol{r})}{\partial N}\right)_{v(\mathbf{r})}$

potential with respect to a perturbation in the external potential. One distinguishes between left- and right-hand derivatives of the electron density with respect to N:

$$f^{+} = \left(\frac{\partial \rho(\boldsymbol{r})}{\partial N}\right)^{+}_{v(\mathbf{r})} \approx \rho_{N+1}(\boldsymbol{r}) - \rho_{N}(\boldsymbol{r}) \tag{60}$$

$$f^{-} = \left(\frac{\partial \rho(\boldsymbol{r})}{\partial N}\right)^{-}_{v(\mathbf{r})} \approx \rho_{N}(\boldsymbol{r}) - \rho_{N-1}(\boldsymbol{r}), \tag{61}$$

where f^{+} describes a nucleophilic and f^{-} an electrophilic attack. Both can be approximated by a finite differences formula, expressing them as differences of the electron density of the system in the initial state and after the addition or subtraction of an electron (at the geometry of the neutral system). Other approaches for the calculation of the Fukui function can be found in [133, 134].

Because the chemical reactivity descriptors (especially the Fukui function) are useful tools in the analysis of chemical reactions also for those mediated by heavy metal containing catalysts [135], it is important to investigate the influence of relativistic effects on them. For the Fukui function, i.e., for the reactivity toward nucleophilic and electrophilic attacks, such a study was carried out by De, Krishnamurty, and Pal [136], who investigated two gold clusters (Au_{19} and Au_{20}), focusing on relativistic effects on vibrational frequencies and on the Fukui function. The authors used effective core potentials (ECPs) to replace the core electrons of the gold atoms in both the relativistic and the nonrelativistic calculations. They find that the incorporation of relativistic effects results in minor changes of the Fukui function.

A second, more detailed study by Sablon et al. [137] investigates the effect of relativistic Hamiltonians on the Fukui function and presents a comparison of four-component Dirac–Coulomb, scalar-relativistic ZORA, spin–orbit ZORA, ECP, and nonrelativistic calculations. $PbCl_2$, Bi_2H_4, and $(CH_3)_2SAuCl$ were chosen as model systems. In contrast to the first study of De et al. [136], Sablon and coworkers report changes in the reactivity, when incorporating relativistic effects in molecules other than gold compounds. The scalar-relativistic calculations and the ECP calculations describe the model systems equally well. The two-component spin–orbit ZORA Hamiltonian and the four-component DC Hamiltonian yield similar results. Comparing these to the scalar-relativistic results shows only minor variations.

From the dataset which is used in [137], Fig. 7 was created, which depicts the Fukui function $f^-(\boldsymbol{r})$ (reactivity toward an electrophilic attack) for the 0.0004 a.u. iso-density surface of the three above-mentioned model systems. The results

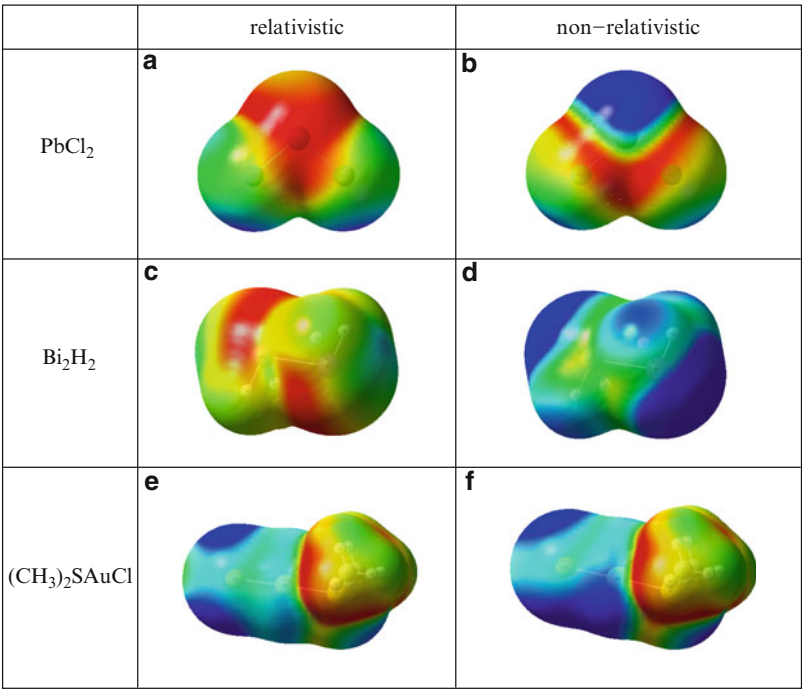


Fig. 7 The Fukui function $f^-(\boldsymbol{r})$ mapped on the 0.0004 a.u. iso-density surface. The relativistic results have been obtained with the spin–orbit relativistic ZORA approach for $PbCl_2$, the four-component methodology for Bi_2H_2, and the scalar-relativistic ZORA approximation for $(CH_3)_2AuCl$. The *color scales* are chosen in such a way that *red* ($f^-(\boldsymbol{r})$ values not higher than 1.0×10^{-4} a.u. for (**a**) and (**b**), 3.0×10^{-5} a.u. for (**c**) and (**d**), and 5.0×10^{-5} a.u. for (**e**) and (**f**) indicates regions of poor reactivity toward electrophilic attacks, whereas *blue* [$f^-(\boldsymbol{r})$ values not lower than 2.0×10^{-4} a.u. for (**a**) and (**b**), 1.5×10^{-4} a.u. for (**b**) and (**c**) and 1.7×10^{-4}] corresponds to highly reactive zones. (Figure created from data presented on [137])

obtained from two-component spin–orbit ZORA ($PbCl_2$), four-component Dirac (Bi_2H_4), scalar-relativistic ZORA [$(CH_3)_2SAuCl$], and nonrelativistic calculations (all model systems) are shown in this Fig. 7. In the case of $PbCl_2$, the major changes in the Fukui function when using a relativistic Hamiltonian are observed directly at the Pb atom, where the relativistic effects are expected to be most pronounced. The nonrelativistic results predict the Pb atom to be the preferred position for an electrophilic attack, whereas the spin–orbit ZORA calculations indicated the opposite, i.e., an attack at one of the two chlorine atoms. For the case of Bi_2H_4, the four-component Dirac calculations predict a high reactivity at both bismuth atoms on the opposite side of the hydrogen atoms. Regarding the nonrelativistic results, the reactivity on the whole isosurface is significantly lowered. Only in the case of the gold complex, the scalar-relativistic result exhibits only minor changes at the gold and the chlorine atoms when comparing to the nonrelativistic calculations.

6.2 The Quantum Theory of Atoms in Molecules

Bader's atoms in molecules (AIM) theory [1, 138] is an interpretative theory which divides the electron density in a molecule into atomic basins $\{\Omega_i\}$ such that it is possible to define properties $G(\Omega_i)$ of an atom i in a molecule, which can then be determined as integrals over the property density $\rho_G(\boldsymbol{r})$ of the corresponding basin:

$$G(\Omega_i) = \int_{\Omega} \mathrm{d}^3 r \rho_{\mathrm{G}}(\boldsymbol{r}). \tag{62}$$

These "atomic" properties are then additive and sum up to molecular properties if all basins are taken into account. In the limiting case of a single atom, the AIM property must be equal to the corresponding property of the free, unbound atom. The individual atoms in a molecule are in this context separated from each other by the interatomic surfaces (IAS), which are usually called zero-flux surfaces. At each point of the IAS, the normal vector $\boldsymbol{n}(\boldsymbol{r})$ is orthogonal to the gradient of the electron density, which is expressed through the so-called zero-flux boundary condition

$$\boldsymbol{n}(\boldsymbol{r}) \cdot \nabla \rho(\boldsymbol{r}) = 0. \tag{63}$$

AIM was first formulated by Bader [1] as a nonrelativistic theory and then later studied by Cioslowski and Karwowski [139] within the relativistic regime. In the nonrelativistic case, the partitioning of the molecule is uniquely defined in terms of open quantum systems (atoms) through the zero-flux boundary condition, which follows from the properties of the Lagrangian density. Due to a certain arbitrariness in the expression for the relativistic Lagrangian density, the partitioning of the atoms is no longer uniquely defined in four-component theory [139].

7 Electron Density in Embedding Schemes

Usually, quantum chemical calculations focus on the calculation of molecular properties like the electron density for a single molecule. Experimental electron densities, however, are obtained from a molecular crystal, in which crystal-packing effects may play a nonnegligible role and have to be considered in the calculation [140–142]. In order to ensure an adequate accuracy in such calculations, the system under investigation might be "embedded" in a suitable environment to include such environment effects [143–148]. This can be achieved by dividing a large molecule of a molecular crystal into smaller pieces such that calculations on each subsystem are feasible. In the optimization of the orbitals of a subsystem, the interaction with the other subsystems of the "environment" can be modeled by an embedding potential. Due to this partitioning, calculations on the full system are avoided, which decreases the computational effort.

One example for such an embedding scheme being a subsystem version of DFT is FDE, which was introduced by Cortona [149] to study properties of solids within DFT. It was then further developed by Wesolowski and Warshel [150], who extended the FDE formalism such that it can also be applied to molecules partitioned into smaller building blocks. The density of the full N-electron system is then partitioned into an "active" subsystem ρ_1, and a second subsystem ρ_2, representing the environment ("frozen" subsystem):

$$\rho_{\text{tot}}(\boldsymbol{r}) = \rho_1(\boldsymbol{r}) + \rho_2(\boldsymbol{r}). \tag{64}$$

The total energy $E_{\text{tot}}[\rho_{\text{tot}}]$, appearing in (53), can then be rewritten as a bifunctional [151] which depends on both subsystem electron densities ρ_1 and ρ_2:

$$\begin{aligned} E_{\text{tot}}[\rho_1, \rho_2] =& E_{\text{NN}} + \int \mathrm{d}^3 r (\rho_1(\boldsymbol{r}) + \rho_2(\boldsymbol{r}))\left(v_1^{\text{nuc}}(\boldsymbol{r}) + v_2^{\text{nuc}}(\boldsymbol{r})\right) \\ &+ \frac{1}{2}\int\int \mathrm{d}^3 r\, \mathrm{d}^3 r' \frac{(\rho_1(\boldsymbol{r}) + \rho_2(\boldsymbol{r}))(\rho_1(\boldsymbol{r}') + \rho_2(\boldsymbol{r}'))}{|\boldsymbol{r} - \boldsymbol{r}'|} + E_{\text{xc}}[\rho_1 + \rho_2] \\ &+ T_s[\rho_1 + \rho_2], \end{aligned} \tag{65}$$

where $v_1^{\text{nuc}}(\boldsymbol{r})$ and $v_2^{\text{nuc}}(\boldsymbol{r})$ are the electrostatic potentials of the nuclei of the subsystems.

Except for a few special cases, the subsystems cannot be expressed in terms of canonical Kohn–Sham orbitals of the full system with the consequence that the kinetic energy cannot be partitioned entirely. There always remains a term which depends on both subsystem electron densities, the so-called nonadditive part of the kinetic energy:

$$T_{\text{s}}^{\text{nadd}}[\rho_1, \rho_2] = T_{\text{s}}[\rho_{\text{tot}}] - T_{\text{s}}[\rho_1] - T_{\text{s}}[\rho_2]. \tag{66}$$

$T_s^{\text{nadd}}[\rho_1, \rho_2]$ is commonly approximated by a kinetic energy density functional. For a given environmental density (e.g., obtained from a KS-DFT calculation of the isolated subsystem), the total energy bifunctional can then be minimized with respect to the active subsystem using the Lagrange method of undetermined multipliers with the constraint that the number of electrons in the subsystem is constant.

The electron density ρ_1 can then be expressed in terms of canonical (subsystem) Kohn–Sham orbitals:

$$\rho_1(\boldsymbol{r}) = 2\sum_{i=1}^{N_1/2} \left|\psi_i^{(1)}(\boldsymbol{r})\right|^2. \tag{67}$$

These orbitals can be evaluated by solving the Kohn–Sham equations with constraint electron density (KSCED):

$$\left[-\frac{1}{2}\nabla_i^2 + v_{\text{eff}}^{\text{KSCED}}[\rho_1, \rho_2](\boldsymbol{r})\right]\psi_i^{(1)}(\boldsymbol{r}) = \varepsilon_i \psi_i^{(1)}(\boldsymbol{r}), \tag{68}$$

where the effective potential $v_{\text{eff}}^{\text{KSCED}}[\rho_1, \rho_2](\boldsymbol{r}) = v_{\text{eff}}^{\text{KS}}[\rho_1](\boldsymbol{r}) + v_{\text{eff}}^{\text{emb}}[\rho_1, \rho_2](\boldsymbol{r})$ is divided into a sum of a KS effective potential and an effective embedding potential which read:

$$\begin{aligned} v_{\text{eff}}^{\text{KS}}[\rho_1](\boldsymbol{r}) &= v_1^{\text{nuc}}(\boldsymbol{r}) + \int \mathrm{d}^3 r' \frac{\rho_1(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}'|} + \frac{\delta E_{\text{xc}}[\rho_1]}{\delta \rho_1} \\ v_{\text{eff}}^{\text{emb}}[\rho_1, \rho_2](\boldsymbol{r}) &= v_2^{\text{nuc}}(\boldsymbol{r}) + \int \mathrm{d}^3 r' \frac{\rho_2(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + \frac{\delta E_{\text{xc}}^{\text{nadd}}[\rho_1, \rho_2]}{\delta \rho_1} + \frac{\delta T_s^{\text{nadd}}[\rho_1, \rho_2]}{\delta \rho_1}. \end{aligned} \tag{69}$$

The FDE approach is in principle equivalent to KS-DFT, given that the exact expression for either the nonadditive kinetic energy or the corresponding kinetic energy potential is known. Because both of them are unknown, the accuracy of FDE strongly relies on the choice of the approximate kinetic energy functional. In order to improve on the quality of the obtained electron density, Wesolowski and coworkers [151] proposed the use of so-called freeze-and-thaw cycles, in which the active and the frozen subsystems are exchanged in an iterative fashion.

Since FDE has been proposed, a number of studies investigate the accuracy of FDE in an indirect way, in terms of criteria as for instance interaction energies [151–156], equilibrium geometries [157], solvent effects [158–160], and molecular properties [161] like dipole moments [162–165] and ESR hyperfine coupling constants [166, 167]. A more detailed review on the applications of FDE can be found in [168–172]. However, only very few studies analyze explicitly the electron density and compare them with reference electron densities obtained from KS-DFT calculations on the full model system.

The first studies dealing with embedding electron densities were presented in 1996 by Stefanovich and Truong [173] who chose the $Li^+\cdots H_2O$ complex as a model system and compare the electron deformation density which is defined as

the difference between the total electron density and the sum-of-fragments electron density for both FDE and KS-DFT results. Wesolowski et al. [152] investigated the hydrogen bonded system (FH···NCH) presenting also electron deformation densities. These publications contain only a brief discussion of the electron densities and do not systematically analyze the influence of different factors on its accuracy.

The first more detailed study concerning the accuracy of the electron density obtained from embedding calculations was presented by Kiewisch et al. [174]. This study focuses on a systematical investigation of different factors like the choice of the basis set, the use of supermolecular basis set expansion (including basis functions, located at the positions of the nuclei of the frozen subsystem) for the optimization of the active subsystem, the number of freeze-and-thaw cycles, and the choice of the exchange-correlation functional as well as the kinetic energy functional. In addition, the changes in the topology of the electron densities were taken into account by analyzing the negative Laplacian at the BCP of the model systems, which can, according to Bader [1], be used to characterize the type of a chemical bond. The study focuses on weakly bound systems, namely $H_2O\cdots F^-$, $F-H-F^-$, and an adenine–thymine DNA base pair, in which the subsystems are connected by hydrogen bonds of different strength. The study states that the choice of the kinetic energy functional, which is used for the approximation of $T_s^{\text{nadd}}[\rho_1, \rho_2]$, plays only a minor role, whereas the application of ghost basis function leads to a significant improvement in the accuracy of the electron density. In general, FDE works quite well for such weakly interacting model systems and the accuracy which is reached when applying an adequate supermolecular basis set in combination with the PW91k kinetic energy density functional is sufficient for many practical applications when about five freeze-and-thaw cycles are applied. The major part of the deficiencies of the embedding density is located in the bonding region, which contains the border between the subsystems and arise due to approximation of $T_s^{\text{nadd}}[\rho_1, \rho_2]$.

The scope of this electron density study was then further expanded in [175] to systems containing coordination bonds and ionic bonds, adopting the conditions from the previous study by Kiewisch et al. (PW91k, supermolecular basis set and five freeze-and-thaw cycles). Ammonia borane was chosen as a model system for donor–acceptor bonds, whereas titanium tetrachloride and chromium hexacarbonyl were incorporated to study ionic bonds of different strength. Even though the electron density for ammonia borane is reproduced acceptably in the bonding region between the subsystems, the negative Laplacian at the corresponding BCP exhibits a wrong sign, which is a strong indication that FDE is not able to give an adequate description for the coordination bond, while the magnitude of the density is qualitatively correct. In contrast to ammonia borane, the density and the negative Laplacian obtained for titanium tetrachloride exhibit fairly less deficiencies. Since both fragments are charged and due to well-known peculiarities of the embedding potential at the frozen subsystem, the so-called "electron leak" problem [163, 173], an unphysical charge transfer between the Cl^- fragment to $TiCl_3^+$ subsystem took place, which could be overcome by additional position-dependent correction terms

proposed by Jacob et al. [164]. In the case of chromium hexacarbonyl, which contains even stronger metal–ligand bonds than titanium tetrachloride, recovering the KS-DFT electron density is a challenge for FDE, because π-backdonation becomes an important effect, which leads to significant covalent contributions in the Cr–C bond. FDE, using the PW91k density functional for the approximation of $T_s^{\text{nadd}}[\rho_1, \rho_2]$, fails in the description of carbonyl complexes, because it is not even possible to recover the expected orbital order and the electron density exhibits huge deficiencies. Since still a lot of effort is put into the development of new kinetic energy density functionals [176], these problems might be solved by the next generation of kinetic energy density functionals.

Because ionic bonds play an important role in crystals of inorganic and metal-lorganic species, we shall provide a closer look on the titanium tetrachloride example from [175]. Titanium tetrachloride is a tetrahedral complex with strong ionic interactions between the central metal atom and the ligands. A BP86/TZP optimized structure, using the above-mentioned position-dependent correction, is depicted in Fig. 8a, whereas contour plots of the electron density for the super-molecular and the FDE calculation can be found in Fig. 8b, c. The major part of the

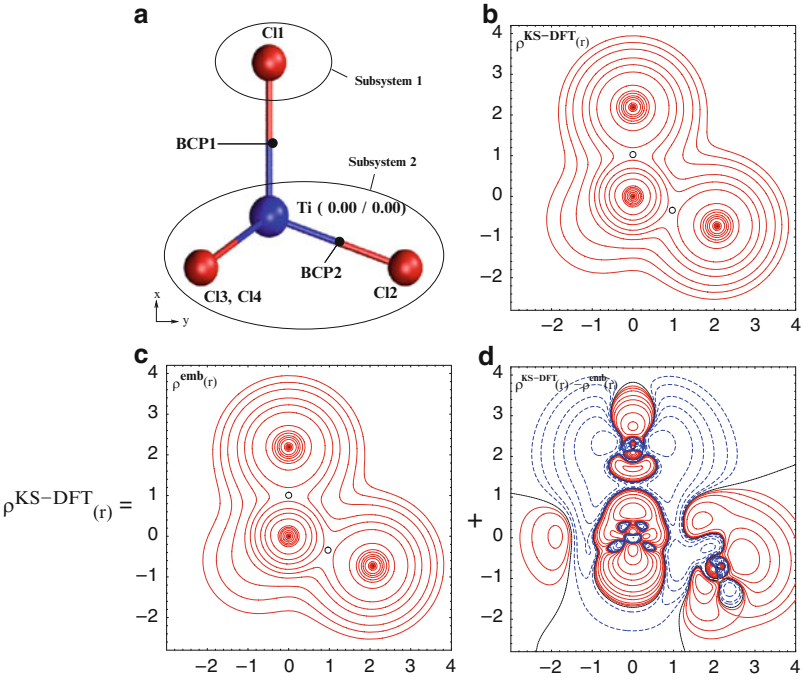


Fig. 8 (**a**) BP86/TZP optimized structure of titanium tetrachloride ($TiCl_4$). The double labeling of some atoms means that there are two atoms that differ only in their z-coordinate and therefore overlay in the picture, (**b**) supermolecular KS-DFT density, (**c**) embedding density, (**d**) difference density $\rho_{\text{super}} - \rho_{\text{emb}}$. (The figure is reprinted with permission from [175]. Copyright 2010 Elsevier)

Table 2 Values of $\rho(\boldsymbol{r})$ in $e\text{Å}^{-3}$ and $L(\boldsymbol{r})$ in $e\text{Å}^{-5}$ at the bond critical point BCP1 of titanium tetrachloride

BCP1	$\rho(\boldsymbol{r})$	$L(\boldsymbol{r})$
Supermolecular	0.65	−1.73
Embedding	0.62	−1.98
Difference	0.03	0.25

deficiencies of the embedding density has its origin in the complex formation, where electron density is shifted from the Cl^- fragment toward the titanium atom. These changes are not fully recovered in the region of the border between both subsystems. The largest changes are observed between both subsystems directly next to the $TiCl_3^+$ fragment on the axis perpendicular to the bonding axis. The difference density between the KS-DFT reference calculation of the full system and the embedding density is shown in Fig. 8e.

Considering the topology of the electron density of titanium tetrachloride, four BCPs are found of which two are located in the plane that is depicted in Fig. 8. BCP1 is located directly next to the border between the subsystems. The two BCPs in the plane are located at the same positions in both the supermolecular density and the embedding density. The values of the electron density and the negative Laplacian at the BCP1, taken from Table 2, differ only by ~0.03 and 0.25 $e\text{Å}^{-5}$, respectively. In contrast to the result for ammonia borane, the negative Laplacian has the correct sign at both BCPs.

8 Conclusion

In this work, we reviewed important issues related to the calculation of the electron density in quantum theory. We discussed the definition of the quantities "electron density" and "current density" as defined by a continuity equation depending on the many-electron Hamiltonian and wave function. In practice, both types of densities are calculated from (quasi)relativistic Hamiltonians and wave function approximations. For this reason, we discussed the most important Hamiltonians – namely the relativistic Dirac–Coulomb–Breit reference, the Dirac–Coulomb, the Douglas–Kroll–Hess, the ZORA and the nonrelativistic Schrödinger Hamiltonians – and wave functions.

In an analysis of approximate electron densities, the most important observations to make regarding the accuracy of approximate Hamiltonians are that in most cases scalar-relativistic variants of the approximate Hamiltonian operators as ZORA and DKH are sufficient to obtain a reasonable description of the electron density. The differences to two-component Hamiltonians including spin–orbit effects are negligible as long as one considers quantities that do not primarily feature spin–orbit effects. These approximate Hamiltonian operators are in general reliable but may produce densities deviating from the fully relativistic reference at the position of the nucleus. For the innermost atomic shells, ZORA and picture-change-affected DKH

densities show significant errors. In the case of the DKH density, these can be overcome by considering a picture-change transformation of the density operator.

Even for interpretive means – such as conceptual DFT – relativistic effects play a significant role as highlighted for the prediction of an electrophilic attack to heavy metal containing molecules and complexes. Although the changes in the Fukui function f^- for the lead and the bismuth complexes are significant, one can hardly detect any differences between relativistic and nonrelativistic Fukui functions in the case of the gold complex.

Finally, we emphasized that FDE is a helpful tool in the analysis of systems, which are embedded in an environment, as is the case for a molecule in a crystal or a molecule in a solvent. We reviewed results for a $TiCl_4$ complex, which was divided into a positively charged $TiCl_3^+$ and a negatively charged Cl^- fragment to show that with FDE even for subsystems connected by an ionic bond, reasonable results for the electron density can be obtained.

9 Computational Methodology

All Hartree–Fock, CISD, CASSCF, and DFT calculations presented in Sect. 4.2 were performed with the MOLPRO2009.1 [177] quantum chemical software package in combination with the second-order DKH Hamiltonian [31, 44, 45] and the atomic natural orbital basis sets (ANO-RCC) [178–180] for which the second set of polarization functions was omitted. In the DFT calculations, the BP86 [181, 182] density functional was used to approximate the exchange-correlation energy. Owing to the large size of the basis set, a counter-poise correction was not considered. The electron densities were calculated with MOLPRO (not including the picture-change effects) and obtained on a cubic grid with step-size 0.02 Bohr (not optimized for relativistic calculations) and then visualized with MATHEMACIA [183]. The coordinates of the linear $Fe(NO)^{2+}$ was obtained from a BP86/TZVPP geometry optimization with TURBOMOLE [184, 185], whereas the coordinates for the bent $Fe(NO)^{2+}$ complex were taken as a subset from a larger complex from [186].

Acknowledgments This work has been financially supported by ETH Zurich (Grant TH-26 07-3). We are grateful to N. Sablon for providing the dataset of the Fukui function for $PbCl_2$, Bi_2H_4, and $(CH_3)_2SAuCl$, which were used to create Fig. 7.

References

1. Bader R (1990) Atoms in molecules. Clarendon, Oxford
2. Matta CF, Boyd RJ (2007) The quantum theory of atoms in molecules. Wiley-VCH, Weinheim
3. Tafipolsky M, Scherer W, Ofele K, Artus G, Pedersen B, Herrmann WA, McGrady GS (2002) J Am Chem Soc 124:5865–5880

4. Scherer W, Sirsch P, Shorokhov D, Tafipolsky M, McGrady GS, Gullo E (2004) Chem Eur J 9:6057–6070
5. Scherer W, McGrady GS (2004) Angew Chem Int Ed 43:1782–1806
6. Friedrich W, Knipping P, von Laue M (1912) Sitz ber Bayer Akademie d Wiss:303–322
7. Bragg WL (1913) Proc R Soc Lond A89:248–277
8. Dickinson RG, Raymond AL (1923) J Am Chem Soc 45:22–29
9. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (1958) Nature 181:662–666
10. Koritsanszky TS (2001) Chem Rev 101:1583–1627
11. Hauptman HA, Karle J (1954) Solution of the phase problem. 1. The centrosymmetric crystal. ACA monograph no. 3, Wilmington
12. Hauptman HA, Karle J (1956) Acta Cryst 9:635–651
13. Patterson AL (1934) Phys Rev 46:372–376
14. Patterson AL (1935) Z Krist 90:517–542
15. Gilmore CJ (1996) Acta Cryst A52:561–589
16. Hohenberg P, Kohn W (1964) Phys Rev 136:864–871
17. Kohn W, Sham LJ (1965) Phys Rev 140:1133–1138
18. Thomas LH (1927) Proc Camb Philos Soc 23:542
19. Fermi E (1928) Z Phys 48:73–79
20. Dirac PAM (1928) Proc R Soc Lond A 117:610–624
21. Dirac PAM (1928) Proc R Soc Lond A 118:351–361
22. Reiher M, Wolf A (2009) Relativistic quantum chemistry. Wiley-VCH, Weinheim
23. Born M (1926) Z Phys 37:863–867
24. Born M, Oppenheimer R (1927) Ann Phys (Berlin) 389:457–484
25. Eckart C (1935) Phys Rev 47:552–558
26. Born M, Huang K (1956) Dynamical theory of crystal lattices. Oxford University Press, New York
27. Szabo A, Ostlund NS (1996) Modern quantum chemistry: introduction to advanced electronic structure theory. Dover, New York
28. Gaunt JA (1929) Proc R Soc Lond A 122:513–532
29. Gaunt JA (1929) Philos Trans R Soc (Lond) A288:151–196
30. Brown GE, Ravenhall DG (1951) Proc R Soc Lond A 208:552–559
31. Hess BA (1986) Phys Rev A 33:3742–3748
32. Reiher M, Wolf A (2004) J Chem Phys 121:2037–2047
33. Reiher M, Wolf A (2004) J Chem Phys 121:10945–10956
34. van Lenthe E, Baerends EJ, Snijders JG (1993) J Chem Phys 99:4597–4610
35. Barysz M, Sadlej AJ, Snijders JG (1997) Int J Quantum Chem 65:225–239
36. Kedziera D, Barysz M (2003) J Chem Phys 116:2696–2704
37. Kutzelnigg W, Liu WJ (2005) J Chem Phys 123:241102
38. Filatov M (2006) J Chem Phys 125:107101
39. Ilias M, Saue T (2007) J Chem Phys 126:064102
40. Sikkerna J, Visscher L, Saue T (2009) J Chem Phys 131:124116
41. Aquilante F, Vico LD, Ferre N, Ghigo G, Malmqvist P-A, Neograd y P, Pedersen TB, Pitonak M, Reiher M, Roos B-O, Serrano-Andres L, Urban M, Veryazov V, Lindh R (2010) J Comput Chem 31:224–247
42. Amsterdam Density Functional program. http://www.scm.com. Theoretical Chemistry, Vrije Universiteit, Amsterdam
43. te Velde G, Bickelhaupt FM, Baerends EJ, Fonseca Guerra C, van Gisbergen SJA, Snijders JG, Ziegler T (2001) J Comput Chem 22:931–967
44. Douglas M, Kroll NM (1974) Ann Phys 1:82, 89–155
45. Wolf A, Reiher M, Hess BA (2002) J Chem Phys 117:9215–9226
46. Reiher M (2006) Theor Chem Acc 116:241–252
47. Wolf A, Reiher M (2006) J Chem Phys 124:064102

48. Wolf A, Reiher M (2006) J Chem Phys 124:064103
49. Reiher M, Wolf A (2007) Phys Lett A 360:603–607
50. Baerends EJ, Schwarz WHE, Schwerdtfeger P, Snijders JG (1990) J Phys B At Mol Opt Phys 23:3225–3240
51. Kello V, Sadlej AJ (1998) Int J Quantum Chem 68:159–174
52. Bučinský L, Biskupič S, Jayatilaka D (2010) J Chem Phys 133:174125
53. Heully JL, Lindgren I, Lindroth E, Lundqvist S, Mårtensson-Pendrill AM (1986) J Phys B At Mol Phys 19:2799–2815
54. Durand P (1986) C R Acad Sci II 303:119
55. Chang C, Pelissier M, Durand P (1986) Phys Scr 34:394–404
56. van Lenthe E, Baerends EJ, Snijders JG (1994) J Chem Phys 101:9783–9792
57. van Leeuwen R, van Lenthe E, Baerends EJ, Snijders JG (1994) J Chem Phys 101:1272–1281
58. Sadlej AJ, Snijders JG, van Lenthe E, Baerends EJ (1995) J Chem Phys 102:1758–1766
59. van Leeuwen R, van Lenthe E, Baerends EJ, Snijders JG (1996) Int J Quantum Chem 57:1272–1281
60. Foldy LL, Wouthuysen SA (1950) Phys Rev 78:29–36
61. Dyall KG, van Lenthe E (1999) J Chem Phys 111:1366–1372
62. Slater JC (1929) Phys Rev 34:1293–1322
63. Condon EU (1930) Phys Rev 36:1121–1133
64. Eickerling G, Mastalerz R, Herz V, Scherer W, Himmel H-J, Reiher M (2007) J Chem Theory Comput 3(6):2182–2197
65. Reiher M (2007) Faraday Discuss 135:97–124
66. Mastalerz R, Lindh R, Reiher M (2008) Chem Phys Lett 465:157–164
67. Rajagopal AK, Callaway J (1973) Phys Rev B 7:1912–1919
68. Hartree DR (1929) Proc Camb Philos Soc 25:225–236
69. White HE (1931) Phys Rev 38:512–520
70. White HE (1934) Introduction to atomic spectra. McGraw-Hill, New York
71. Burke VM, Grant IP (1967) Proc Phys Soc 90:297–314
72. Trautwein A, Harris FE, Freeman AJ, Desclaux JP (1975) Phys Rev B 11:4101–4105
73. Mallow JV, Freeman AJ, Desclaux JP (1975) Bull Am Phys Soc 20:293
74. Mallow JV, Freeman AJ, Desclaux JP (1976) Phys Rev B 13:1884–1892
75. Pyykkö P, Desclaux J-P (1979) Acc Chem Res 12:276–281
76. Ros P, Snijders JG, Ziegler T (1980) Chem Phys Lett 69:297–300
77. Autschbach J, Schwarz WHE (2000) Theor Chem Acc 104:82–88
78. van Wüllen C, Michauk C (2005) J Chem Phys 123:204113
79. Bader RFW, Chandra AK (1968) Can J Chem 46:953–966
80. Baerends EJ, Gritsenko OV, van Leeuwen R (1996) Electron correlation and the structure of the exchange–correlation potential and the correlation energy density in density functional theory. In: Tsipis CA (ed) New methods in quantum theory. Kluwer Academic, Dordrecht
81. Smith VH (1977) Phys Scr 15:147–162
82. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1983) Mol Phys 49:65–89
83. Ritchie JP, King HF, Young WS (1986) J Chem Phys 85:5175–5182
84. Moszynski R, Szalewicz K (1987) J Phys B 20:4347–4364
85. Gatti C, MacDougall PJ, Bader RFW (1988) J Chem Phys 88:3792–3804
86. Gauss J (1991) J Mol Struct 234:95–126
87. He Y, Gräfenstein J, Kraka E, Cremer D (2000) Mol Struct 98:1639–1658
88. Cremer D (2001) Mol Phys 99:1899–1940
89. Jankowski K, Nowakowski K, Grabowski I, Wasilewski J (2009) J Chem Phys 130:164102
90. Jankowski K, Nowakowski K, Grabowski I, Wasilewski J (2010) Theor Chem Acc 125:433–444
91. Vosko SH, Wilk L, Nusair M (1980) Can J Phys 58:1200–1211
92. Lee CT, Yang WT, Parr RG (1988) Phys Rev B 37:785–789

93. Bartlett RJ, Grabowski I, Hirata S, Ivanov S (2005) J Chem Phys 122:034104
94. Boguslawski K, Jacob CR, Reiher M (2010, submitted)
95. Ehrenfest P (1922) Nature 109:745–746
96. Bohr N (1922) Nature 109:746
97. Rosenthal JE, Breit G (1932) Phys Rev 41:459–470
98. Breit G (1958) Rev Mod Phys 30:3408–3411
99. Serafin MM, Peebles SA, Dewberry CT, Etchison KC, Grubbs GS II, Powoski RA, Cooke SA (2004) Chem Phys 298:33–37
100. Serafin MM, Peebles SA, Dewberry CT, Etchison KC, Grubbs GS II, Powoski RA, Cooke SA (2007) Chem Phys Lett 449:33–37
101. Kurian R, Filatov M (2010) Phys Chem Chem Phys 12:2758–2762
102. Kistner OC, Sunyar AW (1960) Phys Rev Lett 4:412–415
103. Shirley DA (1964) Rev Mod Phys 36:339–351
104. Filatov M (2009) Coord Chem Rev 253:594–605
105. Andrae D, Reiher M, Hinze J (2000) Phys Chem Lett 320:457–468
106. Kato T (1957) Commun Pure Appl Math 10:151–177
107. Andrae D (2000) Phys Rep 336:414–525
108. Fricke B, Waber JT (1972) Phys Rev B 5:3445–3449
109. Mastalerz R, Widmark P-O, Roos BO, Lindh R, Reiher M (2010) J Chem Phys 133:144111
110. Visscher L, Dyall KG (1997) At Data Nucl Data 67:207–224
111. Hofstadter R (1956) Rev Mod Chem 28:214–254
112. Svane A, Antoncik E (1987) Phys Rev B 35:4611–4624
113. Knecht S, Fux S, van Meer R, Visscher L, Reiher M, Saue T (2011) Theor Chem Acc (submitted)
114. Filatov M (2007) J Chem Phys 127:084101
115. Hinze J, Jaffe HH (1962) J Am Chem Soc 84:540–546
116. Hinze J, Whitehead MA, Jaffe HH (1963) J Am Chem Soc 85:148–154
117. Hinze J, Jaffe HH (1963) Can J Chem 41:1315–1328
118. Hinze J, Jaffe HH (1963) J Phys Chem 67:1501–1506
119. Pearson RG (1963) J Am Chem Soc 85:3533–3539
120. Pearson RG (1966) Science 151:172–177
121. Pearson RG (1968) J Chem Educ 45:581–587
122. Pearson RG (1968) J Chem Educ 45:643–648
123. Pearson RG (1997) Chemical hardness. Wiley-VCH, New York
124. Fukui K, Yonezawa T, Shingu H (1952) J Chem Phys 20:722–725
125. Fukui K, Yonezawa T, Nagata C, Shingu H (1954) J Chem Phys 22:1433–1442
126. Fukui K, Yonezawa T, Nagata C (1957) J Chem Phys 27:1247–1259
127. Houk KN, Sims J, Duke RE, Strozier RW, George JK (1973) J Am Chem Soc 95:7287–7301
128. Houk KN (1975) Acc Chem Res 8:361–369
129. Fukui K (1982) Angew Chem Int Ed 21:801–809
130. Parr RG, Donnelly RA, Levy M, Palke WE (1978) J Chem Phys 68:3801–3807
131. Geerlings P, Proft FD, Langenaeker W (2003) Chem Rev 103:1793–1873
132. Ayers PW, Anderson JSM, Bartolotti LJ (2005) Int J Quantum Chem 101:520–534
133. Sablon N, Proft FD, Ayers PW, Geerlings P (2007) J Chem Phys 126:224108
134. Sablon N, Proft FD, Geerlings P (2009) J Chem Theory Comput 5:1245–1253
135. Yang W, Parr RG (1985) Proc Natl Acad Sci USA 82:6723–6726
136. De HS, Krishnamurty S, Pal S (2009) J Phys Chem C 113:7101–7106
137. Sablon N, Mastalerz R, de Proft F, Geerlings P, Reiher M (2010) Theor Chem Acc 127:195–202
138. Popelier P (2000) Atoms in molecules: an introduction. Pearson Education, Essex
139. Cioslowski J, Karwowski J (2001) Quantum-mechanical theory of atoms in molecules: a relativistic formulation. In: Carbó-Dorca R, Gironés X, Mezey PG (eds) Fundamentals of molecular similarity. Kluwer Academic, New York

140. Gatti C, Saunders VR, Roetti C (1994) J Chem Phys 101:10686–10696
141. Mladenovic M, Arnone M, Fink RF, Engels B (2009) J Phys Chem B 113:5072–5082
142. Engels B, Schmidt TC, Gatti C, Schirmeister T, Fink RF (2010) Challenging problems in charge density determination: polar bonds and influence of the environment. In: Stalke D (ed) Electron density and chemical bonding. Springer-Verlag GmbH, Heidelberg
143. Govind N, Wang YA, da Silva AJR, Carter EA (1998) Chem Phys Lett 295:129–134
144. Govind N, Wang YA, Carter EA (1999) J Chem Phys 110:7677–7688
145. Klüener T, Govind N, Wang YA, Carter EA (2001) Phys Rev Lett 86:5954–5957
146. Klüener T, Govind N, Wang YA, Carter EA (2002) J Chem Phys 116:42–54
147. Huang P, Carter EA (2006) J Chem Phys 125:084102
148. Iannuzzi M, Kirchner B, Hutter J (2006) Chem Phys Lett 421:16–20
149. Cortona P (1991) Phys Rev B 44:8454
150. Wesolowski TA, Warshel A (1993) J Phys Chem 97:8050
151. Wesolowski TA, Weber J (1996) Chem Phys Lett 248:71–76
152. Wesolowski TA, Chermette H, Weber J (1996) J Chem Phys 105:9182–9190
153. Wesolowski TA (1997) J Chem Phys 106:8516–8526
154. Wesolowski TA, Ellinger Y, Weber J (1998) J Chem Phys 108:6078
155. Kevorkyants R, Dulak M, Wesolowski TA (2006) J Chem Phys 124:024104
156. Neugebauer J (2008) J Phys Chem B 112:2207–2217
157. Dulak M, Kaminski JW, Wesolowski TA (2007) J Chem Theory Comput 3:735–745
158. Neugebauer J, Louwerse MJ, Baerends EJ, Wesolowski TA (2005) J Chem Phys 122:094115
159. Neugebauer J, Jacob CR, Wesolowski TA, Baerends EJ (2005) J Phys Chem A 109:7805–7814
160. Jacob CR, Neugebauer J, Jensen L, Visscher L (2006) Phys Chem Chem Phys 8:2349–2359
161. Neugebauer J (2009) J Chem Phys 131:084104
162. Jacob CR, Wesolowski TA, Visscher L (2005) J Chem Phys 123:174104
163. Dulak M, Wesolowski TA (2006) J Chem Phys 124:164101
164. Jacob CR, Beyhan M, Visscher L (2007) J Chem Phys 126:234116
165. Jacob CR, Visscher L (2008) J Chem Phys 128:155102
166. Wesolowski TA (1999) Chem Phys Lett 311:87–92
167. Neugebauer J, Louwerse MJ, Belanzoni P, Wesolowski TA, Baerends EJ (2005) J Chem Phys 123:114101
168. Wesolowski TA (2006) One-electron equations for embedded electron density: challenge for theory and practical payoffs in multi-level modeling of complex polyatomic systems. In: Leszczynski J (ed) Computational chemistry: reviews of current trends, vol 10. World Scientific, Singapore
169. Jacob CR, Visscher L (2010) Towards the description of covalent bonds in subsystem density-functional theory. In: Wesolowski TA, Wang YA (eds) Recent advances in orbital-free density functional theory. World Scientific, Singapore (to appear)
170. Neugebauer J (2009) ChemPhysChem 10:3148–3173
171. Neugebauer J (2010) Phys Rep 489:1–87
172. Neugebauer J (2010) Orbital-free embedding calculations of electronic spectra. In: Wesolowski TA, Wang YA (eds) Recent advances in orbital-free density functional theory. World Scientific, Singapore (to appear)
173. Stefanovich EV, Truong TN (1996) J Chem Phys 104:2946–2955
174. Kiewisch K, Eickerling G, Reiher M, Neugebauer J (2008) J Chem Phys 128:044114
175. Fux S, Kiewisch K, Jacob CR, Neugebauer J, Reiher M (2008) Chem Phys Lett 461:353–359
176. Wang YA, Carter EA (2000) Orbital-free kinetic-energy density functional theory. In: Schwartz SD (ed) Theoretical methods in condensed phase chemistry. Kluwer, Dordrecht
177. Werner H-J et al (2009) MOLPRO, version 2009.1, a package of ab initio programs. http://www.molpro.net
178. Roos BO, Lindh R, Malmqvist P-A, Veryazov V, Widmark P-O (2005) J Phys Chem A 109:6575–6579

179. Roos BO, Lindh R, Malmqvist P-A, Veryazov V, Widmark P-O (2005) Chem Phys Lett 409:295–299
180. Roos BO, Lindh R, Malmqvist P-A, Veryazov V, Widmark P-O, Borin AC (2008) J Phys Chem A 112:11431–11435
181. Becke AD (1988) Phys Rev A 38:3098–3100
182. Perdew JP (1986) Phys Rev B 33:8822–8824
183. Wolfram Research, Inc (2008) Mathematica version 6.0. Wolfram Research, Inc, Champaign, IL
184. Ahlrichs R, Bär M, Häser M, Horn H, Kömel C (1989) Chem Phys Lett 162:165–169
185. Ahlrichs R et al. Turbomole. http://www.cosmologic.de/turbomole.html
186. Conradie J, Ghosh A (2007) J Phys Chem B 111:12621–12624

Struct Bond (2012) 147: 143–192
DOI:10.1007/430_2010_26
© Springer-Verlag Berlin Heidelberg 2010
Published online: 7 December 2010

Residual Density Analysis

Kathrin Meindl and Julian Henn

Abstract In this chapter the foundations and applications of the Residual Density Analysis (RDA) are shown. The RDA is a concept for the detection and quantification of features from residual density grids. These may be from XD, MoPro, TONTO, or BayMEM. It can be used in radial function development, data processing and data reduction development, in the development of refinement strategies, as a fingerprint method for systematic errors and their imprint onto the residual density, and in day-to-day applications to the density and thermal motion models. In particular, the RDA is used as a stopping criterion in Multipole Modeling as it gives a measure for structural information (features) in the residual density distribution. When no more features are present, thermal motion and density models fit the experimental data with residuals distributed according to a Gaussian and only noise remains in the residual density. The RDA cannot give a proof for the correctness of a model, but it can disprove the expected matching with Gaussian residuals between model and data. Applications of the RDA to electron reconstructions from Multipole Models and from an application of the Maximum Entropy Method are given. Section 3.8 gives an application of the RDA as a fingerprint method.

Keywords Charge density · Systematic sources of error · Topology

Remark The content of this chapter "Residual Density Analysis (RDA)" is based on and expands the ideas in the Dissertation "Residual density validation and the structure of Labyrinthopeptin A2" by K. Meindl [1]. The interested reader is referred to this work, in particular for more examples of the impact of multipole model errors on the residual density distribution as given by an RDA-plot.

K. Meindl
Georg-August Universität Göttingen, Göttingen, Germany
J. Henn (✉)
Universität Bayreuth, Bayreuth, Germany
e-mail: julian.henn@uni-bayreuth.de

Contents

1 Motivation 145
1.1 Special Needs in Charge-Density Studies 145
1.2 The Limited Use of *R*-Values 146
1.3 The Residual Density: Flat and Featureless? 148
1.4 Learning from Mistakes 149
1.5 1001 Hidden Sources of Error 151
1.6 The "Traditional" Approach for Evaluating the Quality of a Refinement 153
2 Basic Ideas and Aims of the RDA 154
2.1 The Fractal Dimension Approach 155
2.2 The Improper Probability Density Approach: e_{gross} and $|\Delta F|$ 162
3 Applications 163
3.1 ... to Idealized Distributions 164
3.2 ... to Experimental Resolution Truncation 166
3.3 ... to Quantify Features in a Model Refined Against Experimental Data 166
3.4 ... to a Multipole Refinement 174
3.5 ... to an Electron Density Reconstruction with the MEM 176
3.6 ... to Extinction Problems 179
3.7 ... to Statistical Disorder 181
3.8 ... to an Anharmonic Motion Problem 182
4 Future Developments 189
4.1 Physical Measures of Quality 189
4.2 Further Tools for Analysis, Advanced Theories, and Adjustments to Charge-Density Needs 190
4.3 A Common Effort 191
References 192

Abbreviations, Terms, and Symbols

$d^f(\rho_0)$	Fractal dimension of residual density value ρ_0. Also fractal dimension distribution of the residual density in the whole unit cell
e_{gross}	Gross residual electrons
e_{net}	Net residual electrons
GC	Gram Charlier coefficients
GoF	Goodness of Fit
IAM	Independent Atom Model
MEM	Maximum Entropy Method
MM	Multipole Model
p.d.f.	Probability density function
POF	Percentage of Features
RDA	Residual Density Analysis
RDA-plot	Fractal dimension distribution; synonyms: graphical representation of $d^f(\rho_0)$ for all values of ρ_0
R_{free}	*R*-value
S	(Information) Entropy
TDS	Thermal Diffuse Scattering
VSCC	Valence Shell Charge Concentration

ρ_0 Residual density, difference density. Synonyms for the density obtained from a Fourier synthesis with Fourier absolute coefficients from the difference of the absolute structure factors and phases from the model

χ Extinction coefficient; also weighted residual sum (χ^2) for least squares fitting

1 Motivation

1.1 Special Needs in Charge-Density Studies

It is an easy task to give the bee-line distance between Göttingen and Berlin to an accuracy of several 10 km; it is a harder task to give the distance between the "Gänseliesl" in Göttingen and the "Brandenburger Tor" in Berlin to the accuracy of several meters, and it is almost impossible to give the distance between the bill of the goose and a detail of the "Brandenburger Tor" like the tip of the nose of Viktoria to the accuracy of a few millimeters.

Also, the possible sources of error are very different for these tasks. While in the first case not even the curvature of earth's surface has to be taken into account, nor differences in altitude, in the second case both will have to be considered. In the third case, one might find that this distance is not even well defined and changes with time due to tidal forces or local oscillations caused by heavy traffic.

A similar situation arises when comparing standard structure determination for small molecules to charge-density studies and finally to the topological analysis with the typical discussion of the second derivatives of the electron density: while standard structure determinations are rather robust and still give chemically meaningful results also when structure parameters are affected by measurement errors and parameter correlation, this need not be the case for charge-density studies, and for the Laplacian it might even be questioned whether this is a well-defined quantity, when static densities are constructed by deconvolution from a time- and space average, the dynamic density.

When aiming at the highest accuracy, the emphasis shifts from avoiding coarse mistakes to measurement, assessment, and control of all possible sources of error. Also, instruments are needed to discriminate between true "measured" properties of the electron density and artifacts from random- or systematic errors. The quotation marks are used, because in reality the electron density is not measured, but reconstructed from the diffraction data. Reconstruction implies use of diffraction-, density-, and thermal motion-models, all of which may contain several simplifications. In the end, there may be chemically and physically different solutions to the density reconstruction problem, which fit the diffraction data equally well.

These considerations are made a bit more explicit in the following paragraphs.

1.2 The Limited Use of R-Values

There are a variety of statistical descriptors and quality measures available in crystallography. The probably most important single concept in use is the comparison or agreement between measured and predicted entities such as intensities or moduli of structure factors. Several definitions are used differing in the entity they refer to (intensities or structure factors) and in the weights assigned to individual terms. These are called *R*-values, where the "*R*" abbreviates "residuals" – the part of the data, that is left over, or which remains undescribed by the model. As the residuals may be positive and negative, the absolute residuals or the residuals squared are considered. These are summed up to the absolute residual sum, giving a positive number or zero.

$$R_1 = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}, \tag{1}$$

$$wR_1 = \sqrt{\frac{\sum_{hkl} w||F_{\text{obs}}| - |F_{\text{calc}}||^2}{\sum_{hkl} w|F_{\text{obs}}|^2}}, \tag{2}$$

$$R_2 = \sqrt{\frac{\sum_{hkl} (I_{\text{obs}} - I_{\text{calc}})^2}{\sum_{hkl} (I_{\text{obs}})^2}}, \tag{3}$$

$$wR_2 = \sqrt{\frac{\sum_{hkl} w(I_{\text{obs}} - I_{\text{calc}})^2}{\sum_{hkl} w(I_{\text{obs}})^2}}. \tag{4}$$

Also, the "Goodness of Fit," *GoF*, and the R_{free} use similar residual sums. One might think of these *R*-values in terms of a distance of the actual model from the experimental data. Perfect agreement would result in each *R*-value being zero. But perfect agreement is not expected, as the experimental data unavoidably contain noise [2]. Furthermore, there may be a large number or even infinite many density models with the same *R*-value (even if the *R*-value is zero!), as the phase information is lost in these measures. And finally, the reconstruction process implies fitting of a model to the data. For the fitting procedure, a choice must be made, which entity (which *R*-value) is used for minimizing the distance between the experimental data and the model. Note that different choices lead to different "best models".

1.2.1 Smaller *R*-Value: Better Model?

It has been emphasized above that the resulting model depends on the entities considered (intensities or structure factors) for the comparison and, furthermore, in which way measured and calculated entities are compared (unit- or other weights).

This already makes clear that there is no strong correspondence between the quality of a model and its R-values, because parameter changes affect different entities such as intensities and structure factor moduli differently, resulting in different "optimum models".

To give an example: let us assume the experimenter decides to fit the density model with respect to the agreement between measured and predicted structure factor moduli. He finally gets a resulting R_1-value of W. Now he wants to compare the quality of these results with other published data, however, these were given as R_2-values. So he calculates what the R_1-value means, when translated to an R_2-value and gets a result X. After reexpressing his results, he has the idea to do the refinement again, this time directly with respect to the intensities, instead of the structure factor moduli and he gets a result $R_2 = Y$. We kindly ask the reader to stop here for some 30 s and think about the following question: is this result Y greater, equal or smaller than X? Please make an educated guess before you continue reading.

Before we come back to this question, we briefly discuss another situation. Suppose there is excellent high-resolution data and an independent atom model (IAM) is fitted to the data. Let us assume the atoms move only harmonically. After the model converged to its final values, anharmonic motion parameter refinement is included for all heavy atoms, which are still considered to be spherically symmetric. The R-values will fall further, however, the resulting density model is not necessarily better, as anharmonic motion parameters will artificially account partly for the aspheric electron density due to chemical bonding, to compensate for the inadequate spherical static density model. In this case, a smaller R-value just expresses that the experimentally and theoretically obtained intensities are in better agreement; however, they are in *better* agreement for the *wrong* reasons. This kind of error, to obtain better agreement for the wrong reason, is very important in charge-density studies. Although this last example may be a bit trivial, however, it expresses one point very clearly: R-values do not say anything about how physically or chemically reasonable a model is. R-values do not *prove* a model to be better than another one, they just indicate agreement but the agreement can be achieved by any combination of density- and thermal motion parameters, which need not be physically and chemically meaningful. This is of special importance in charge-density studies, with highly flexible models, correlating model parameters, and small R-value differences between competing models.

Now back to the other example. Did you make a guess? You still can, if you do not read on. But now we have to go on and solve the riddle. The R_2-value obtained for a refinement against structure factor moduli was X, while the R_2-value for a refinement against intensities was Y. Of course, Y will in general be smaller than X. That is because it makes a difference when refining against $|F|$ or against I and an optimum for one case is not necessarily also an optimum for the other case.

When the experimenter now decides to compute the R_1-value from $R_2 = Y$, he will obtain a result Z that is in general larger than W.

1.2.2 Global and Local Measures

Despite the shortcomings of *R*-values already mentioned, there is yet another aspect of quality measures that becomes increasingly important in charge-density studies. This is the question of local and global measures.

In charge-density studies, typical questions are about the details of the electron density: is there a positive or a negative charge at this atom? Are there VSCCs at a given atom? And so on. Two different charge-density models may have more in common than differences, e.g., when one model employs an Al^{3+} scattering factor, whereas in another model the Al atom is considered to be neutral. Or in one model anharmonic motion of an atom is considered, and in the other not. Or two models just differ in the set of radial screening parameters for the spherical core density κ or for the aspheric valence density κ'.

Because the changes in the models are so small, the *R*-values, which measure the total or *global* agreement, also change only little. It would sometimes be helpful if the *local* description also could be compared, i.e., if the local environment of an atom or a group of atoms could be analyzed with respect to the influence of certain parameters like those above mentioned.

Summary: *R*-values just measure global agreement, but there may be many different density models, which all agree to the same degree with the data, even in the case of vanishing *R*-values. In contrast to standard structure determination, in charge-density studies *R*-values are also not able to establish the superiority of one model over another. As a consequence, further quality criteria in addition to *R*-values are needed.

1.3 The Residual Density: Flat and Featureless?

What is known about the properties of the residual density? In the relevant literature, one finds the phrase that the residual density should be "flat and featureless" [3] and that this is the most important single criterion. But what exactly does this mean? "Flat" is easy to understand, it means that the absolute values of residual density peaks and holes should be as small as possible. But how is the content of features in the residual density assessed? How is it possible to prove the presence of features in the residual density?

To show that the residual density does not contain features is of importance for several reasons. First of all, it proves that the model is able to describe the experimental findings in the whole unit cell and that the model parameter values have been adjusted appropriately. Features in the residual density are a hint toward an incomplete or inappropriate model, wrong parameter values, or hidden sources of error. Similar to the *R*-values, also the absence of features in the residual density does not prove the model to be correct in a mathematical sense, but the opposite holds: features in the residual density prove inappropriateness in the model-to-data relationship. The source may be in the experiment (data measurement, data processing) or in the theoretical and modeling part (inappropriate density- or thermal motion

model, oversimplified diffraction theory). We will come to the possible sources of error later more explicitly. The most important thing here is: as long as the residual density contains features, there is information in the residual density about a mismatch between data and model. This information can be used to identify and eliminate sources of error, which makes the final results more accurate and reliable. For this, however, it must be clear how to measure features in the residual density.

> The residual density should be flat and featureless. But how can it be shown that a given residual density is flat and featureless?

1.4 Learning from Mistakes

1.4.1 Software Handling Errors

Although this topic does not seem to be of scientific interest, it is, however, of practical interest, especially for those, who actually do the refinements. As science is not taking place in an ideal world with ideal working conditions but in the real world, with sometimes short notice of deadlines, high expectations of the results and over hours, fixed-term contracts, with changes in the scientific staff, it is only natural, that handling errors occur. The interest here is of course in those handling errors, which are not automatically detected by the software. An example for this in a multipole refinement is given when a multipole model is set up in a way that resets hydrogen bond distances to preassigned values from neutron diffraction data, but accidentally the RESET command is not invoked. The hydrogen atoms are then not reset; however, their positions are also not refined. This introduces an error which might not be detected as errors stemming from hydrogen atoms often have seemingly only a low impact. One could easily go on in the refinement without realizing that something went wrong. When finally, at the end of the whole refinement, the error is detected, it might not be sufficient to just activate the RESET command and much of the work has to be redone. This example might appear a little specialized, however, many more such handling errors exist and are sometimes not detected. If the residual density distribution was concomitantly analyzed, this kind of errors can be avoided. This, however, requires an easy-to-use and easy-to-interpret analysis tool.

1.4.2 Disorder

Another topic of interest is disorder, which might not be obvious, when the molecule under consideration is large and has bulky groups, such as ^{t}Bu, when the occupation number is close to one or when a disordered atom is in a bond. Disorder may, nevertheless, have a significant impact on the model due to parameter correlations.

A spatially well-localized error in the model may influence all other parameters due to parameter restrictions and correlations. As an example, one may think of the monopole parameters. These are interconnected by the restriction that they sum up to the total number of electrons, which is constant during the refinement. If disorder is present but not taken into account, the monopole values (and with high probability other model parameters, too) of the disordered part are wrong. This also affects the other monopole values *via* the above-mentioned constraint. If disorder is present, it needs to be described. The problem may be to detect disorder in the first place.

1.4.3 Overfitting

The number of model parameters and observed data needs to be in balance. For many parameters and few observations, there may be many parameter value combinations, which fit the experimental data all very well. Not a single set of parameter values out of these needs to be physically or chemically meaningful. A problem of this kind would arise, for instance, when the diffraction data are not of the resolution needed for a charge-density study, however, a multipole expansion is performed. Despite this inappropriateness of the model, the R-values would of course decrease. This may sound trivial, but it is not, because high-resolution data may be sufficient for the refinement of dipoles of the heavy atoms, however, for quadrupoles and higher moments, the data may not be sufficient. Where exactly is this border between overfitting and appropriate parameterization? Also in these cases, it would be helpful to analyze the residual density locally and globally. If the local analysis does not indicate progress in the fitting, the atom is probably overfitted. Analysis tools may help in these cases also to mark the border between *predicted* density parameter values such as the ones obtained from an Invariom refinement [4] and parameter values *derived* from the data.

1.4.4 Sharing Experiences

In the day-to-day work of charge-density refinements, the search for systematic errors has to be done over and over again for each individual refinement. There are individual errors, which will probably repeat and there are errors more specific to the charge-density problem at hand. How much easier would it be, if there was an indicator warning of certain mistakes or systematic errors. It would be helpful if one could picture the effect of an error and would know immediately, what is going on, because the characteristic imprint of this error was described earlier in the literature. One would learn faster not only from own mistakes, but also from those made by others. The refinement process would become more efficient. Moreover, one would learn the more, the more mistakes are made. The word “mistake” should be taken here without valuation, as a neutral word describing a probably unexpected mismatch or discrepancy between predicted and measured values or between predicted

and reconstructed charge-density distributions. The charge-density community would take advantage of such indicators as a whole, because everybody would learn from everybody else, if those indicators existed and if the charge-density community as a whole was open enough to discuss such "critical" things.

1.5 1001 Hidden Sources of Error

Some sources of error were already given above. What else may cause a mismatch between predicted and measured entities? This list might get quite long. The reason is that the more accurate the experiment gets the finer the effects are that disturb or even destroy the results – sometimes without a warning and with no obvious sign. Whereas standard structure determination is rather robust and may forgive small systematic errors, the situation differs in charge-density studies. Two examples are given in the following.

Suppose there are Renninger [5] reflections in the measured data. In standard structure determination, these play a role mainly by preventing the determination of the correct space group, as systematic absences are obstructed by multiple diffracted intensities. Once the correct space group has been established, the problem of Renninger reflections can in most cases safely be ignored, although these diffract not only in otherwise systematic absent reflections but in others, too. Also, the missing intensity from the strong reflection that is multiply diffracted does not play a crucial role. In charge-density studies, however, all of these errors will affect the values of model parameters, e.g., the monopole values.

The other example is thermal diffuse scattering (TDS). It is well known that TDS may account for more than 25% of the intensities of the weak reflections [6]. In standard structure determination, this will not affect bond distances, but thermal motion parameters. These are, however, affected only little. In charge-density studies, these artificially large intensities may be described by artificially induced changes in the mono- and multipoles and will certainly affect the anharmonic motion parameters (if refined), which values are determined mainly from the high order data.

Both topics (and many more), Renninger reflections and TDS, will become more and more important in future charge-density studies as for higher and higher accuracy these effects will be more and more disturbing. By now, there may be hundreds of experimental data sets waiting for new methods to correct the data a posteriori or the theory for these effects. More, possibly hidden, sources of error are:

- Errors on operating the diffractometer, the data processing software or the refinement software
- Accidental disturbance of the measurement by temporary power supply breakdown, temperature fluctuations at the detector, the crystal or at the X-ray source, cosmic rays, natural radioactivity
- Errors in absorption and extinction corrections, Renninger reflections

- Errors in Lorentz and polarization corrections
- Errors in the error model
- Thermal diffuse scattering
- Limited flexibility of the density model
- Disorder and libration
- Anharmonic motion
- Signal overflow at the detector
- Electric noise in the detector, dark currents
- Radiation damage of the crystal
- Wrong scattering factors, e.g., nonrelativistic for heavy atoms
- Beam impurities such as $\lambda/2$ scattering, diverging beam, nonflat beam profile or spectral truncation
- Higher order diffraction
- Crystal not bathing in the beam, not adequately centered, moving during measurement, chemically instable, etc.
- Wrong, missing or inappropriate incident beam correction
- Inappropriate conversion factors for the detector
- Data scaling
- Additional diffraction by ice, the pin holding the crystal, oil, air, the beam stop, or the monochromator
- Data integration errors

Admittedly, some effects have been mentioned several times now, however, it is also obvious that the list is not complete. Some of the systematic errors should be attributed to the theory. As an example, let us take again the Renninger reflections. Their appearance is a natural consequence of a very high crystal quality and strong intensities and is expected. However, the simplified theory used does not include the description of Renninger reflections, therefore the standard theory applied, which provides the connection between the electron density distribution in the unit cell and the diffraction pattern, should be extended to include these effects. This, however, is not accomplished on the fly.

One may optimistically assume that extinction corrections are small for small extinction coefficients $\chi << 1$ (for a definition of χ, see Equation (26)). It has been shown, however, that this is not true [1]. For an extinction coefficient of $\chi = 0.049$, the "correction" is much larger than all changes induced by employing a multipole model instead of the spherical IAM. As a consequence, little errors or only a little imprecision in the correction method might have a large effect on the density model. We will come back to this example later.

If one tries to minimize a systematic error by changing the experimental conditions, one almost always reinforces another error. For example, the lower the experimental temperature, the better rotation disorder of methyl groups is prevented; however, at the same time the probability for extinction effects rises due to the lower thermal motion and accordingly better diffraction.

Also, some of the systematic errors will compensate each other at least partly. A hypothetical example: when the background signal is measured too strong, it may

compensate for some reflections with the presence of TDS, which peaks under the Bragg signals. In this case of *error compensation*, the effective deletion of one source of error may deteriorate the situation as long as the counteracting source of systematic errors is not deleted, too.

May this be as it is, the validity of the experimental results is limited by the systematic errors. This situation is in charge-density studies much more demanding than in conventional crystallography.

Any one of the above-mentioned possible sources of error has the potential to completely spoil a charge-density study. This might not even be obvious, as small *R*-values and good fitting statistics may give the impression that everything is fine.

It is therefore of increasing importance to assess, control, and quantify these errors as the wind needs to be tempered to the shorn lamb.

1.6 The "Traditional" Approach for Evaluating the Quality of a Refinement

Until recently, it was common to "prove" that the residual density is flat and featureless by showing difference density plots in selected planes, often chosen in a way as to contain the heaviest atoms. Sometimes, these plots were calculated for a resolution well below of the experimental one, insinuating explicitly, or, more frequently, implicitly, that features appearing at higher resolutions are pure artifacts from experimental noise – which is not the case. The selection of the planes and the step width of the contour lines were chosen freely. If the plane containing the highest difference peak and the deepest hole, which also often appear in the vicinity of the heaviest atom, was not shown, usually nobody would ask for it.

This situation has come about since there was a common understanding that difference densities need to be featureless, but no generally accepted standard procedure how to prove that a difference density is without features was known. There were just no general standards available. It is, however, well known that the largest peak and deepest hole need not appear in a molecular plane, and that features tend to accumulate at special positions or symmetry axes in the unit cell. For this reason alone, it is important to analyze the whole unit cell and not only subjectively selected planes. Furthermore, there should be a criterion available, which tells whether features in the residual density are of truly statistical random error. If they are not, this means that the model is inadequate and all properties derived from the model may be spoiled. It is even more problematic that it is impossible to tell which of the model-derived properties are not affected. But as indicated in the introduction, as a rule of thumb one can regard standard structure parameters as being relatively robust, whereas the second derivatives of the charge-density and the Laplacian are regarded as most sensitive. The Laplacian, however, is most important when the electron density is interpreted chemically, as it is connected to the valence shell charge concentrations (VSCCs).

2 Basic Ideas and Aims of the RDA

It would go much too far to state that the RDA is able to solve all the above-mentioned problems, but it is true that the RDA touches these in one way or the other. For example, the RDA allows for an assessment of the whole residual density in the unit cell. The RDA plots allow for an estimate to which degree maximum peak and deepest hole fit naturally in the residual density distribution. How "flat" and "featureless" the residual density is can be seen from an RDA plot. The difference between an MM or a MEM refinement and an IAM refinement can be quantified in terms of "gross residual electrons" or in "percentage of features" (POF). The influence of data processing such as extinction correction can be translated to a number of gross residual electrons. This number may be compared to the difference in gross residual electrons from IAM and MM or IAM and MEM, thereby giving an impression of how sensitive the output of an MM or MEM refinement is with respect to extinction correction procedures. In summary, the RDA is a first step toward unifying and standardizing the description of experimental (and data processing) sources of error.

The basic idea is that the residual density contains extremely valuable information about modeling errors and modeling inadequacies. Modeling in this context comprises also data collection and reduction as well as the scattering theory, the density- and the thermal motion models, and all theories, assumptions, approaches, and simplifications entering at any stage, e.g., assumptions about the background signals or spot profiles. Why is data collection and reduction a part of modeling? Because the data collected is supposed to be Bragg data corresponding to monochromatic X-rays deflected from a crystal with perfect periodicity.

The basis idea is further to extract this information from the residual density, i.e., to analyze the residual density, hence Residual Density Analysis. In the second step, the information gain should be used to improve the modeling, if necessary.

A very important aim is also to give the term "featureless" a meaning, i.e., to derive a quality criterion that tells whether or not the residual density distribution in the whole unit cell can be considered featureless. The development of these ideas was continuously accompanied by the question how things would be in an ideal world. This sharpens the expectations and makes it easy to differentiate between the general and the special cases. So, for deriving the residual density descriptors, we just follow this line and start with answering the question:

How would the residual density look like after a refinement in an ideal world? In a very ideal world without noise and background signals and with a perfect diffractometer, with a perfect detector and a perfect scattering theory, a perfect imperfect crystal and unlimited resolution, there are no residuals at all and correspondingly no peaks and holes in the residual density. If you try to picture this residual density distribution in the unit cell in your mind, you will see a completely flat "distribution" with only one value, which is zero and which completely fills the unit cell. Mathematically, this situation can be described

either as a probability density distribution with probability one for the residual density value zero and with probability zero for other residual density values. It can also be described as a three-dimensional manifold of the residual density value zero completely filling the unit cell. For the next paragraph, we will stick to the second meaning.

What follows from this picture when there are small imperfections in the experiment, say, a little noise in the data? The resulting residual density distribution cannot be a three-dimensional manifold any more, because there will be holes in this distribution with positive as well as negative residual density values, separated by the zero residual density manifold. The zero residual density manifold will topologically reduce to a surface that is wrapped into the unit cell. This zero residual density surface, when unwrapped, may comprise a very large surface, much larger than any unit cell surface. If this area approaches infinity, which will happen when the surface is bent forth and back on every spatial scale, then the three-dimensional manifold is approached again. On the other hand, if this surface becomes approximately as large as a unit cell surface (in an order of magnitude sense), the topological surface also approaches a two-dimensional manifold. In all other cases, a manifold with a dimensionality between 2 and 3 is approached.

Which value is approached exactly can be measured in the framework of fractals, fractal dimensions, or, more precisely, fractal dimensionality [7]. The word "fractal" is Greek and means "broken." This refers to the dimensionality of the manifolds, which, in contrast to the Euclidian, topological dimensions we are used to (zero, one, two, and three) may, but need not, assume "broken," that is noninteger values.

Of course, the residual density value zero is just an example, the dimensionality can be determined for all values between the deepest residual density hole, $\rho_{0,\min}$, and the largest residual density peak, $\rho_{0,\max}$. As a final result, one gets the fractal dimension distribution, $d^{\mathrm{f}}(\rho_0)$ (more accurately it would be termed fractal dimensionality distribution, however, the idea was coined a "fractal dimension" concept and that is the term to be used here), of the residual density in the whole unit cell.

2.1 The Fractal Dimension Approach

2.1.1 Why Fractals?

Should the reader search for a deeper rooting of the RDA in the area of fractals, he or she will sooner or later be disappointed. There is no deeper connection of these topics as the connection already mentioned: For an ideal model, and ideal, noise-free data, and unlimited resolution, the zero residual density manifold will be a three-dimensional continuous manifold completely filling the unit cell. Any deviation from this ideal situation therefore must result in a manifold with dimensionality smaller than three. That is it!

The fractal dimension approach was just the access that was chronologically chosen first. Other approaches also exist. For example, in Sect. 3.2 the improper

probability density approach will be discussed. Both approaches are completely equivalent.

2.1.2 The Residual Density Descriptors e_{gross}, e_{net}, and $d^{\mathrm{f}}(\rho_0)$

The residual density descriptors should be able to describe the residual density distribution conclusively. Therefore, these descriptors should give answers to the questions: What is the total error? Is there an excess or a deprivation of electrons in a part of the unit cell? Is the residual density flat? Is the residual density featureless?

2.1.3 The Total Error and e_{gross}

The total error, including the noise, is described by the number of "gross residual electrons":

$$e_{\mathrm{gross}} = \frac{1}{2} \int_V |\rho_0(\mathbf{r})| \mathrm{d}^3 r. \tag{5}$$

When every electron is in its correct place, only noise is left over and the number of gross residual electrons will be at a minimum value. This measure may, but need not, be evaluated for the whole unit cell. If evaluated for the whole unit cell, the number of gross residual electrons is a *global* measure of quality. If evaluated for a part of the unit cell, e.g., in a box around the heaviest atom, it is a *local* measure.

2.1.4 Local Excess Electrons and e_{net}

The total number of electrons in the unit cell is described by F_{000}, which corresponds to the undeflected beam that is not measured and to the constant term in the Fourier sum. By scaling, it is assured that $\Delta F_{000} = F_{000,\mathrm{obs}} - F_{000,\mathrm{calc}} = 0$, such that by definition there will never be excess electrons or a lack of electrons in the whole unit cell. When, however, only a part of the unit cell is considered, the number of "net residual electrons," e_{net}, may assume a positive or a negative number.

$$e_{\mathrm{net}} = \int_V \rho_0(\mathbf{r}) \mathrm{d}^3 r. \tag{6}$$

While e_{gross} is always a positive number, e_{net} may be positive or negative or zero. From the definitions follows

$$-e_{\text{gross}} \leq \frac{1}{2} e_{\text{net}} \leq e_{\text{gross}}. \tag{7}$$

When one of the equal signs holds, all errors in the volume under consideration stem from missing or excess electrons.

2.1.5 Features in the Residual Density and $d^{\text{f}}(\rho_0)$

First, it is explained how the fractal dimension is defined and then how it is implemented.

For a given manifold to calculate its fractal dimension, one divides the manifold in cuboids (or in quadrangles) of characteristic length ε. Then the number N of cuboids is counted, which contain the part of the structure of interest. In this application, this structure would be a residual density iso-value ρ_0. The fractal dimension of this residual density iso-value is then defined as:

$$d^{\text{f}}(\rho_0; \varepsilon) = \frac{\log(N(\varepsilon))}{\log(1/\varepsilon)} \tag{8}$$

with

$$1/\varepsilon = \sqrt[3]{3(n_x - 1)\,(n_y - 1)\,(n_z - 1)} \tag{9}$$

and n_x, n_y, n_z being the number of residual density grid values in x, y, and z direction.

The determination of $N(\varepsilon)$ is important. The residual density grid mesh is assumed to be so fine that all values between two adjacent residual density grid points are taken only once.

Although this concept can be used to quantify features more precisely, as will be shown later, the most important information is just drawn from a graphical representation of the fractal dimension for all residual density values: If the shape is a parabola, the residuals are distributed according to a Gaussian. The residuals are not systematic then and there is no structural information contained in it. On the other hand, if there are shoulders, if the distribution is asymmetric or not a parabola, this clearly indicates that the residual density contains information for improvement of the refinement.

For the importance of the logarithmic scale, see Figs. 1 and 2. These refer to the same data, however, presented in different ways. A residual density grid was calculated for the multipole refinement of experimental high-resolution data of $S(N^tBu)_3$ [8]. In Fig. 1, the frequency of residual density values has been counted

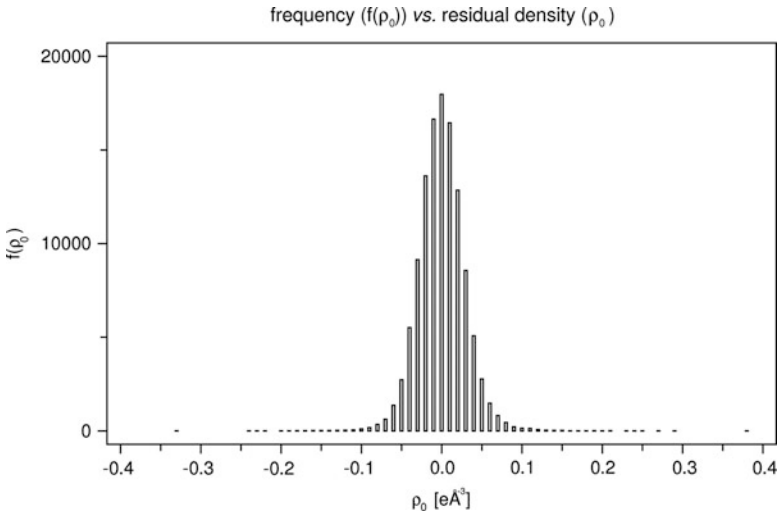


Fig. 1 Frequency f of residual density values ρ_0 after multipole refinement of $S(N^tBu)_3$. Is it a Gaussian, yes or no?

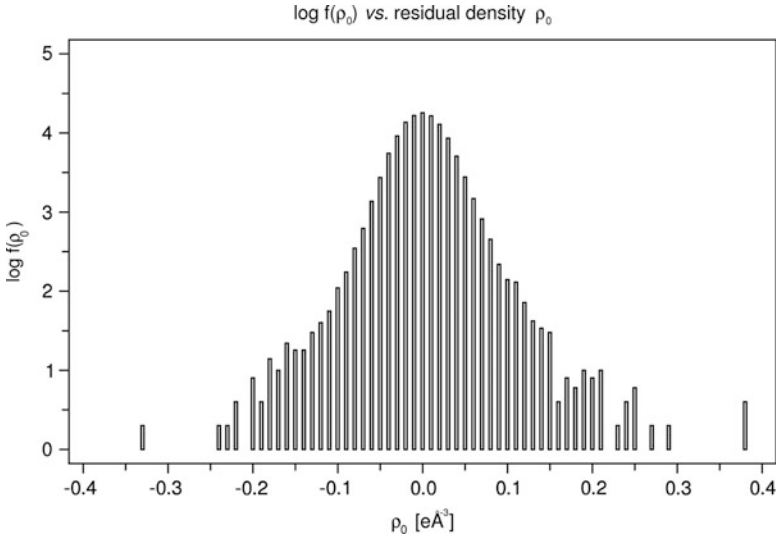


Fig. 2 Logarithm of the frequency f of residual density values ρ_0 after multipole refinement of $S(N^tBu)_3$. If the residuals were distributed according to a Gaussian, the shape of this logarithmic representation would be a parabola

and represented in form of a histogram. The reader should now have a look at the figure and try to estimate, whether or not the distribution of the histogram shows a Gaussian curve, before proceeding with reading.

Many readers will hopefully agree that the shape is clearly reminiscent of a Gaussian: there is only one maximum, the distribution is symmetric, and there is the typical change in curvature. Only in the periphery, the symmetry is not maintained strongly and there are small gaps, but this may as well be so just by accident.

The logarithm of the frequency distribution is shown in Fig. 2. In the range close to the center, it resembles a parabola, however, in the periphery the deviations from the parabolic shape are quite distinct and they do not give the impression of random statistical fluctuations but of being systematically too large. This is indeed the case.

While Figs. 1 and 2 demonstrate the importance of the logarithmic scale, there are still two important steps missing to convert Fig. 2 into an RDA plot. The difference in the RDA is *what* is counted and *how* it is counted. Let us start with the latter. Whereas in Fig. 1 and 2 it was counted how many grid values are within a given small range, e.g., between 0.10 and 0.11 $e\text{Å}^{-3}$, i.e., the data were binned, in the RDA it is counted which residual density values are between neighboring residual density grid points, i.e., the data are not binned.

For example, when at adjacent grid points, say in x-direction, the residual density values are -0.1682 and $-0.0976\ e\text{Å}^{-3}$ then in the frequency table the values for all integer multiples of $0.01\ e\text{Å}^{-3}$ between these limits are incremented by one. That is for the residual density values -0.16, -0.15, -0.14, -0.13, -0.12, -0.11, $-0.10\ e\text{Å}^{-3}$. This is done for each of the three independent directions. So far for the "how", now for the "what".

While in the histogram approach small grid volumes are counted (i.e., grid points with attached volume, with a residual density value within certain limits), in the RDA the faces of grid points containing certain residual density values (such as all positive and negative integer multiples of $0.01\ e\text{Å}^{-3}$) are counted. For an example, see Fig. 3. The blue line represents the residual density value zero as an example

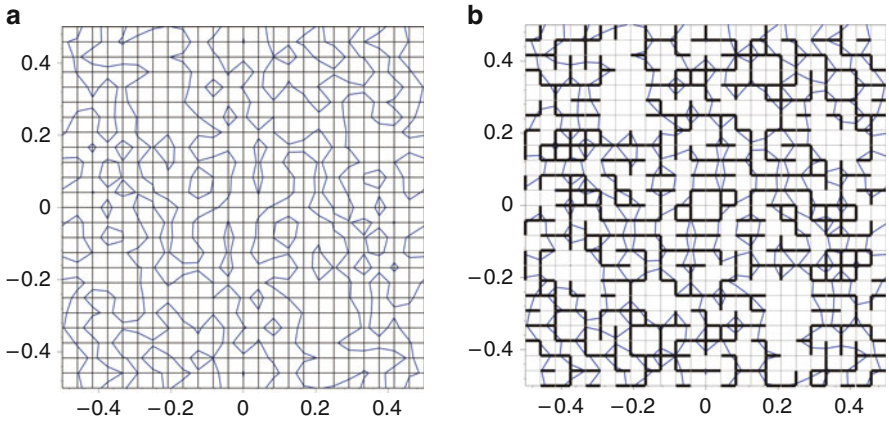


Fig. 3 (**a**) Residual density value zero (*blue curved line*) and residual density grid (*straight gray lines*) of an arbitrary example structure. (**b**) Evaluation of a residual density plane. The short black lines contribute to $d^{\mathrm{f}}(\rho_0 = 0)$

and the gray lines represent the grid (Fig. 3a). The short black lines in Fig. 3b indicate those gridlines, which are intersected by the residual density value zero. For the fractal dimension determination of the residual density value zero $d^{\mathrm{f}}(\rho_0 = 0)$ only those short black lines contribute. For a three-dimensional grid, the borders between adjacent grid points are no longer lines but surfaces.

Equations (8) and (9) still depend on ε, which is a bit of a drawback when one wants to have an absolute scale. However, for comparisons of different density or thermal motion models fitted to the same experimental data, this is already sufficient.

The dependence on ε raises the question: is there a natural choice for the residual density grid? On the one hand, the grid should be fine enough to show all relevant details, however, it also needs to be in balance with the experimental resolution.

2.1.6 Resolution in Real and in Reciprocal Space

When diffraction data are collected up to a given resolution and the collected data are Fourier-transformed to restore the original function, the high frequency components above the resolution threshold are missing, one has effectively applied a low-pass filter. Therefore, the restored function shows less detail, it is smoother. Accordingly, if this restored function is sampled in real space, there is a minimum threshold for the spacing, which, when used, will enable to exactly reproduce the Fourier coefficients. If the spacing is made finer, only computing power and memory is wasted.

For example, an audio CD is sampled with a rate of approximately 42 kHz. This is sufficient to restore all frequencies in the music up to 21 kHz, which is assumed to be satisfactory for most people.

The connection between sampling frequency and the maximum frequency reconstructible, the bandwidth limit, is given by the Nyquist sampling theorem [9]. This theorem imposes restrictions on the choice of n_x, n_y, and n_z in (8) and (9). We are now going to have a look what the Nyquist sampling theorem implies for the RDA.

Suppose the data were measured completely up to a resolution now given in the maximum indices h, k, and l. From the Nyquist sampling theorem, we can immediately tell the spatial resolution to be chosen for the residual density grid.

$$\begin{aligned} n_x &= 2|h|_{\max} \\ n_y &= 2|k|_{\max}. \\ n_z &= 2|l|_{\max} \end{aligned} \tag{10}$$

2.1.7 Expectation Value for $d^{\mathrm{f}}(0)$

If we take (10) as a convention, we can put it into (8) and (9) for the calculation of an expectation value for $d^{\mathrm{f}}(\rho_0 = 0)$. How many times can the residual density value zero be assumed in a residual density grid?

There are $(n_y - 1)(n_z - 1)$ faces of cuboids in one x-layer and $(n_x - 1)$ of those layers, which we chose to match $(2|h|_{\max} - 1)$, resulting in the appearance of the residual density value zero to be at most $(2|h|_{\max} - 1)(n_y - 1)(n_z - 1)$ for the x-direction. Similar formulas can be derived for the y- and z-directions, leading in total to

$$\begin{aligned} N_{\max}(\rho_0 = 0) = {} & (2|h|_{\max} - 1)\,(n_y - 1)\,(n_z - 1) \\ & + (2|k|_{\max} - 1)\,(n_z - 1)\,(n_x - 1) \\ & + (2|l|_{\max} - 1)\,(n_x - 1)\,(n_y - 1) \end{aligned} \tag{11}$$

which is the maximum number. We are, however, interested in the average number. This is derived from (11) by taking the half, because every time a border is crossed between two adjacent residual density grid values, there is a chance of ½ that the residual density sign changes. If it changes, the residual density value zero has been assumed at least once somewhere between the grid points (although the value itself need not appear as a grid value).

$$\langle N(\rho_0 = 0)\rangle = \frac{1}{2} N_{\max}(\rho_0 = 0). \tag{12}$$

Taking (8)–(12) together, it is possible to calculate the expectation value of $d^{\mathrm{f}}(0)$ with a pocket calculator from the maximum indices in h, k, and l. For example, for $|h|_{\max} = |k|_{\max} = |l|_{\max} = 25$ one obtains:

$$\frac{\log(0.5 \times 3 \times 49^3)}{\frac{1}{3}\log(3 \times 49^3)} = 2.83721. \tag{13}$$

To test the derived formula, a grid of $50 \times 50 \times 50$ Gaussian distributed random numbers was generated 20 times with mean value zero. The grid was analyzed with the Residual Density Analysis software jnk2RDA [10], which additionally to the numbers e_{gross} and e_{net} gives $d^{\mathrm{f}}(0)$ and a plot of $d^{\mathrm{f}}(\rho_0)$. The mean value and standard deviation were: $d^{\mathrm{f}}(0) = 2.83723 \pm 0.00041$, which is in good agreement with the predicted value from (13).

2.1.8 Quantification of Features: Percentage of Features

The results from the last paragraph can be employed to quantify the amount of features. It was shown that the value $d^{\mathrm{f}}(0)$ scatters, although only little, around its mean value. Therefore, the ratio of the actual value and the expected value may give values slightly over 100%, which is counterintuitive.

It is therefore suggested to use the following definition as a measure of the POF in the residual density distribution:

$$\mathrm{POF} = 100 \left| 1 - \frac{d^{\mathrm{f}}(0)}{\langle d^{\mathrm{f}}(0)\rangle} \right|, \tag{14}$$

with the choice of the grid to be analyzed in accordance with (10). Applied to the 20 random grids from above, the POF gives at most 0.034% features and the minimum value is 0.001%. Up to one digit, there are 0.0% features in these grids which is expected for random grids. For an application to experimental data, see Sect. 3.

2.2 *The Improper Probability Density Approach:* $e_{\mathbf{gross}}$ **and** $|\Delta F|$

As mentioned earlier, there is another way to perceive the fractal distribution of the residual density. Suppose it is known how the residual density is distributed in the unit cell from a probability or frequency point of view, i.e., it is known how many times each residual density value appears. This is a one-dimensional function and therefore of less information content than the whole three-dimensional residual density distribution in the unit cell. If this function is normalized, i.e., if the area of the histogram or under the curve is chosen to equal unity by an appropriate choice of the normalization constant, the function may serve as a probability density function (p.d.f.).

Now, it is important to realize that this reduced knowledge is already sufficient to calculate all properties we are interested in via the usual formulas of statistical expectation values. As an example, we take e_{gross} and define it now from the statistical point of view rather than in (5). This leads to

$$e_{\text{gross}} = \frac{1}{2} V \langle |\rho_0(\mathbf{r})| \rangle \tag{15}$$

and

$$\langle |\rho_0| \rangle = \int_{-\infty}^{\infty} p(\rho_0) |\rho_0| \mathrm{d}\rho_0. \tag{16}$$

If the p.d.f. $p(\rho_0)$ is known, it can be substituted into (16) and the result into (15) for the calculation of e_{gross}.

There is a nice application of (15) and (16) for the limiting case we are most interested in, that is when the refinement is at or close to the optimum and the p.d.f. is a Gaussian:

$$p(\rho_0) dp = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{\rho_0^2}{2\sigma^2}} dp. \tag{17}$$

Substituted in (16), this yields after integration:

$$\langle |\rho_0| \rangle = \sqrt{\frac{2}{\pi}} \sigma. \tag{18}$$

On the other hand, also the second moment can be calculated from the p.d.f.,

$$\left\langle |\rho_0|^2 \right\rangle = \int_{-\infty}^{\infty} p(\rho_0)|\rho_0|^2 \mathrm{d}\rho_0, \tag{19}$$

which yields in the special case of (17)

$$\left\langle |\rho_0|^2 \right\rangle = \sigma^2. \tag{20}$$

From (18) and (20) follows [1]:

$$\langle |\rho_0| \rangle = \sqrt{\frac{2}{\pi}\left\langle |\rho_0|^2 \right\rangle}. \tag{21}$$

There is another way to calculate the mean value of the modulus squared of a function whose Fourier series is known by means of Parseval's theorem:

$$\left\langle |\rho_0|^2 \right\rangle = \sum_{hkl} |\Delta F|^2, \tag{22}$$

which leads, when substituted into (15) and (21), to the handy formula:

$$e_{\mathrm{gross}}\big|_{V_{\mathrm{UC}}} = \sqrt{\frac{2}{\pi}\sum_{hkl} |\Delta F|^2}. \tag{23}$$

Equation (23) gives the connection between the numbers of gross residual electrons (i.e., the total error including noise) in the whole unit cell and the structure factor differences. Equation (23) is valid in the case of a Gaussian p.d.f. of the residual density. Please note that this measure is identical to $wR_1|_{w=1}$ [i.e., wR_1, where the weights have been chosen to be unit weights, see (2)], if the factor of proportion is chosen accordingly. Unit weights are in accordance with the expansion of a function in an orthogonal and normalized basis set like in a Fourier series. However, while wR_1 is a global measure, e_{gross} can also be evaluated locally on a residual density grid.

This result gives a different view on the definition of the fractal dimension of the residual density distribution. The RDA plot can also be interpreted as the logarithm of an improper (not normalized to unity) probability density function.

3 Applications . . .

In the following, some examples of applications of the RDA are given. As these examples represent different states of development of the theory, they are not at the same stage. For example, in Sec. 2.2 "the improper probability density approach:

e_{gross} and |ΔF|", it has been shown that e_{gross} can be evaluated from the absolute modulus of the difference structure amplitudes for the case that the residual density obeys a Gaussian distribution. Obviously, this number can be compared to the number obtained from summing over the grid. Equality of these numbers is another, independent indicator of a Gaussian distribution of the residuals, additional to the parabolic shape and to the expectation value for $d^f(0)$.

For the evaluation of the residual density grids, the software jnk2RDA [10] was used throughout. It accepts residual density grids from XD, MoPro [11, 12], TONTO [13], and BayMEM [14]. For BayMEM, the grid format has to be modified before evaluation. Output data is a Postscript file with an RDA-plot, i.e., a plot $d^f(\rho_0)$ versus ρ_0 for a range -1.0 to $+1.0$ eÅ^{-3}. Additionally, the values for $d^f(0)$, $\rho_{0,min}$, $\rho_{0,max}$, and their difference $\Delta\rho_0 = \rho_{0,max} - \rho_{0,min}$ are given. If the data were in XD format, some additional information from the master file is automatically extracted, like the unit cell data for calculation of the volume. The multipole model used, the grid size of the residual density grid and the XDFOUR section are printed into the Postscript file. From version 1.4 also the entropy S of the absolute values of the residual density distribution is given [for a definition of S see (25)]. This software is available for Unix/Linux and Windows machines.

3.1 ... to Idealized Distributions

To check the basic assumptions and to have a look at how noise appears in the RDA-plot, ideal, noise-free data are generated from a multipole model for $S(N^tBu)_3$ up to a resolution of $\sin\theta/\lambda = 1.14$ Å^{-1}. These data have been added to Gaussian noise with mean value zero according to the following equations:

$$I_{hkl}^{noise} = I_{hkl}^{ideal} + I_{hkl}^{error}$$
$$I_{hkl}^{error} = p_1 \sqrt{I_{hkl}^{ideal}} \text{Random Gauss}(x)$$
$$\text{Random Gauss}\,(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$\sigma = 1$$
$$\mu = 0$$

The noise increases with p_1. In Fig. 4, the RDA-plot is shown for four different values of the noise control parameter p_1: 0.000, 0.222, 0.444, 0.888, and in Table 1 the residual density descriptors are given. Additionally, there is a horizontal and a vertical red bar drawn in Fig. 4. The horizontal bar indicates the flatness (actually only, when drawn at $d^f(\rho_0) = 0$), $\Delta\rho_0 = \rho_{0,max} - \rho_{0,min}$. This value depends explicitly solely on the following four factors:

(1) The experimental resolution
(2) The residual density grid resolution
(3) The density and thermal motion model
(4) The noise in the data

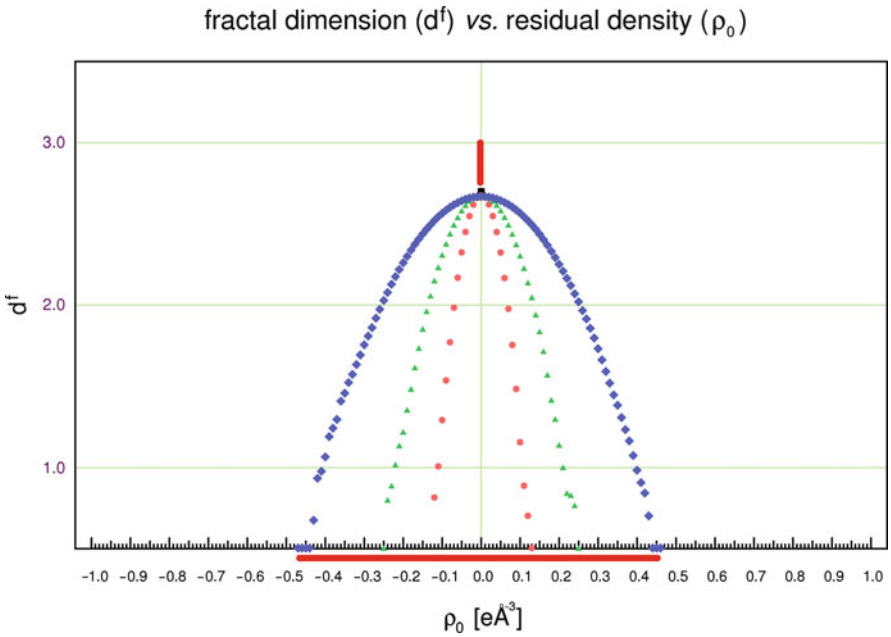


Fig. 4 Ideal parabolic shape for the fractal dimension distribution (RDA-plot) in the presence of Gaussian noise and absence of model errors. For a discussion of the distributions and the *horizontal* and *vertical red bars*, see the text. *Black square*: $p_1 = 0.000$; *red circles*: $p_1 = 0.222$; *green triangles*: $p_1 = 0.444$; *blue rhombuses*: $p_1 = 0.888$

Table 1 Residual density descriptors applied to simulated data on $S(N^tBu)_3$ for differing levels of Gaussian noise as quantified by the noise control parameter p_1

p_1	$d^f(0)$	e_{gross} [e]	$-\rho_{0,min}$ [eÅ^{-3}]	$\rho_{0,max}$ [eÅ^{-3}]	$\Delta\rho_0$ [eÅ^{-3}]
0.000	2.7956	0.34	0.00	0.00	0.00
0.222	2.7693	8.38	0.12	0.12	0.24
0.444	2.7678	16.41	0.22	0.23	0.45
0.888	2.7647	30.92	0.41	0.46	0.87

Of course, there is an implicit dependency on the data processing, too, which is ignored for this discussion. For a fixed experimental resolution and residual density grid, and a perfect model, the baseline is in proportion to the noise. This is exactly what Fig. 4 shows: the more noise, the broader the distribution. The figure also shows deviations from the ideal parabolic shape in the periphery. This just expresses the usual statistical fluctuations, which do not play a role for a very large number of data points (like in the center at $\rho_0 = 0$), but become increasingly dominant with a decreasing number of data points.

The vertical red line, which shows the difference between the actual value of the zero residual density value $d^f(0)$ and the ideal value $d^f(0) = 3.0$ depends on the

same factors (1)–(4) given above. The dependency on these factors is discussed in detail in [10]. Additionally, this value $d^f(\rho_0)$ is related to the features in the residual density, as has been shown in Sect. 2.1.8.

The RDA-plot shows the formerly separated properties "flatness" and "features" of the residual density in one graphic. Under ideal circumstances, when no systematic errors are present, this distribution is parabolic in shape.

Table 1 shows an increasing number of gross residual electrons and an increasing difference between maximum peak and deepest hole for increasing noise, whereas the fractal dimension at the residual density value zero simultaneously decreases a little bit, which is expected under the assumption that strong noise also tends to spread spatially wider than weak noise. Even for no noise, $p_1 = 0.000$, e_{gross} remains finite, although one would of course expect it to approach zero. This difference is explained by the observation that in XD the structure factor calculation does not yield identical results between input structure factors and output structure factors, when no refinement is performed, despite using the correct input file for the model parameters. Instead, changes in the last digit occur for a number of structure factors. These little differences add up to a nonzero value for e_{gross}.

3.2 . . . to Experimental Resolution Truncation

As already mentioned above, the experimental resolution is sometimes truncated for the difference density plots. For the determination of the parameters from a least-squares fit, however, the whole data is used. It has also been mentioned that this procedure is not advisable, as there exists valuable information about the adequateness of the model to the data in the high-frequency data. It is indeed that this data reveals best otherwise hidden sources of error. To demonstrate the effect of experimental resolution truncation, again the $S(N^tBu)_3$ data is used. The fully refined model is used for the calculation of $50 \times 50 \times 50$ residual density grids, however, the resolution is truncated from 1.14 to 1.00 and 0.80 Å^{-1}, respectively. The same model parameters are used throughout. Figure 5 shows the RDA-plots for the whole unit cell and contour plots in the molecular plane.

The more the resolution is truncated the smoother and more acceptable the residual density appears. The model, however, has not been changed and the valuable information in the high-frequency data that indicates a nonperfect model and systematic differences between model and data is suppressed by the cutoff.

3.3 . . . to Quantify Features in a Model Refined Against Experimental Data

In this example, POFs of an IAM refinement and of an MM refinement against experimental data of $S(N^tBu)_3$ are calculated. These data have been used for the histograms in Figs. 1 and 2.

Additionally, for a comparison the percentage of features of an ideal data set of structure factors, with the same resolution and a comparable number of gross residual electrons, are calculated, too. The gross residual electrons are then caused

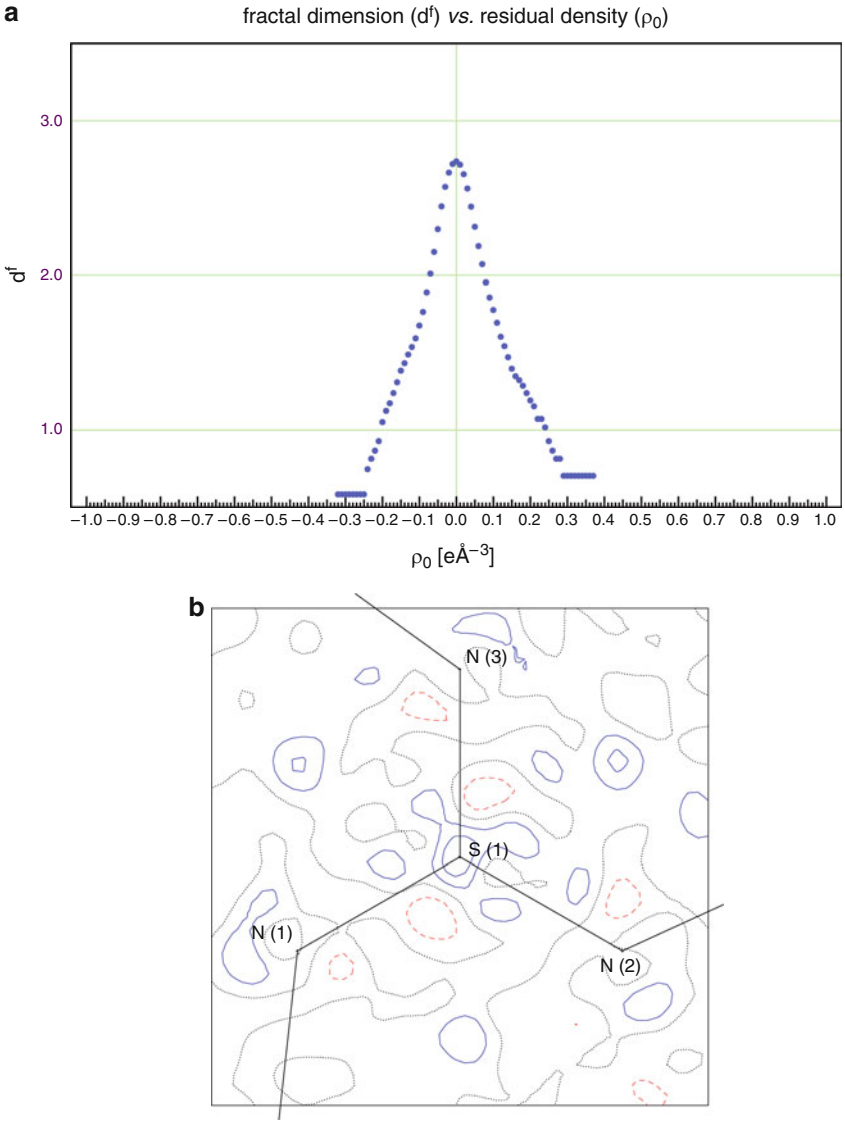


Fig. 5 Effect of the truncation of the experimental resolution on the difference density maps. *Left*: RDA-plots for the whole unit cell. *Right*: Contour plots through the molecular plane. *Blue solid lines*: positive residual density, *red dashed lines*: negative residual density, *gray dotted lines*: zero residual density. The contour spacing is 0.1 e $Å^{-3}$. Resolution 1.14 (**a**, **b**), 1.00 (**c**, **d**), 0.80 (**e**, **f**) $Å^{-1}$

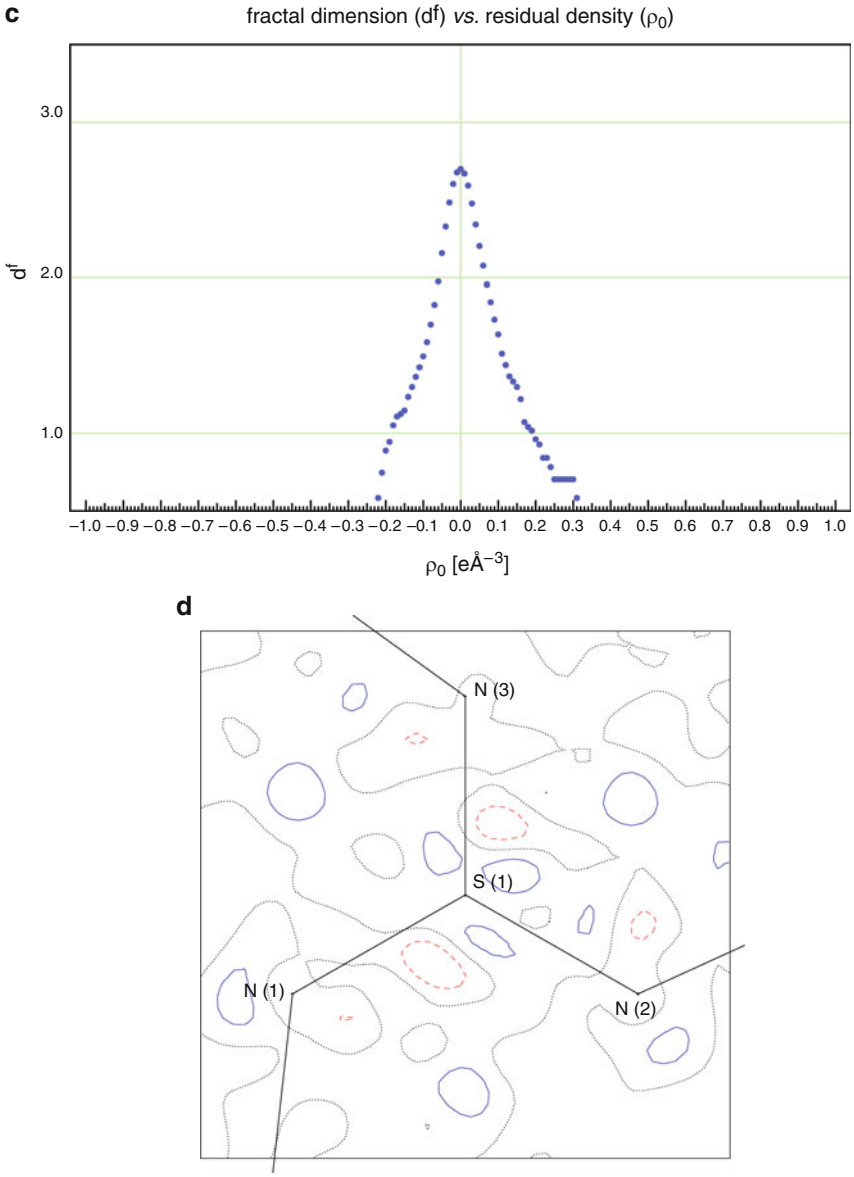


Fig. 5 (continued)

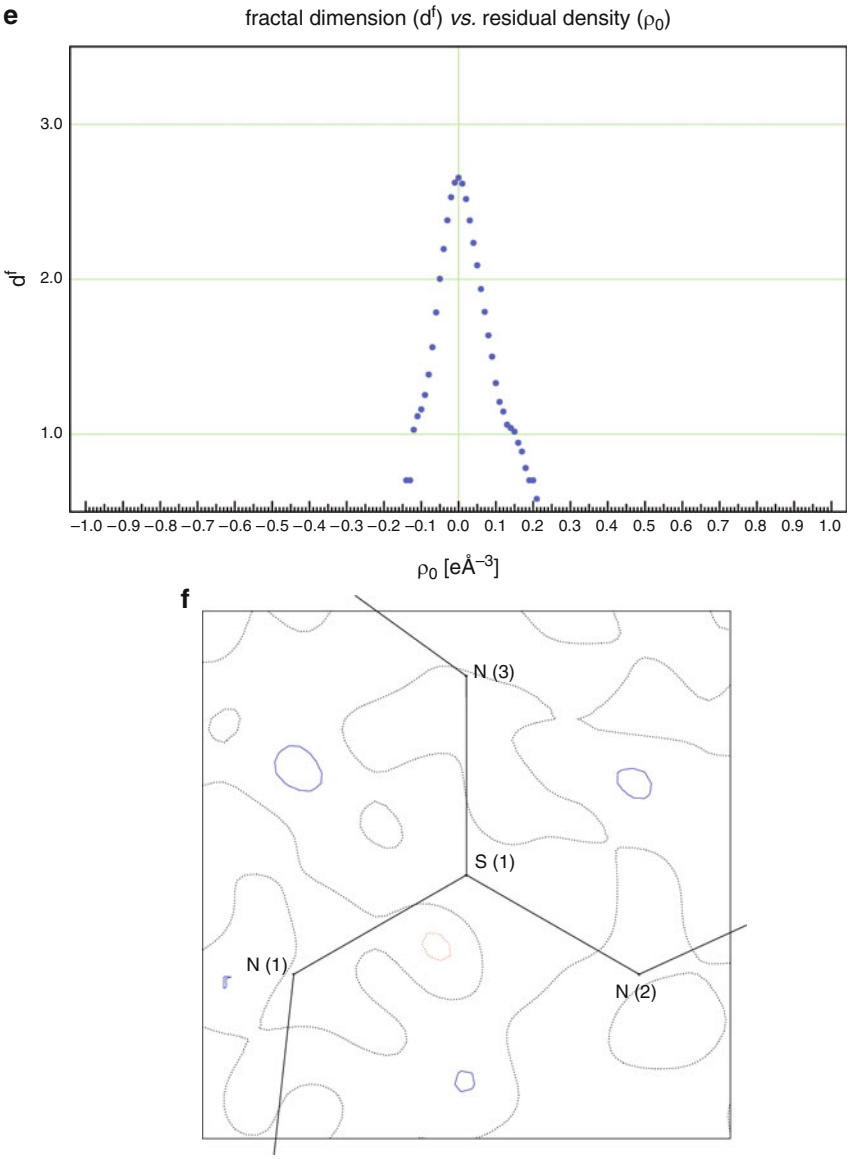


Fig. 5 (continued)

Table 2 Residual density descriptors for identical models at different experimental resolution

Cutoff resolution [Å^{-1}]	$d^f(0)$	e_{gross} [e]	$-\rho_{0,min}$ [eÅ^{-3}]	$\rho_{0,max}$ [eÅ^{-3}]	$\Delta\rho_0$ [eÅ^{-3}]
1.14 (no cutoff)	2.7366	8.39	0.32	0.37	0.69
1.00	2.6989	7.25	0.23	0.32	0.55
0.80	2.6543	6.27	0.15	0.21	0.36

Table 3 Residual density descriptors including entropy S and percentage of features POF for experimental IAM and MM refinements and for a perfect MM refinement on synthetic data with comparable total error e_{gross} and therefore comparable $wR_1|_{w=1}$ values as in the experimental MM refinement

	$d^f(\rho_0)$	e_{gross} [e]	$-\rho_{0,min}$ [eÅ^{-3}]	$\rho_{0,max}$ [eÅ^{-3}]	$\Delta\rho_0$ [eÅ^{-3}]	S	POF [%]
IAM (exp.)	2.6826	11.55	0.73	0.59	1.32	69,564	5.3
MM (exp.)	2.7076	7.21	0.24	0.31	0.55	30,574	4.4
MM (theo.)	2.7786	7.36	0.10	0.11	0.21	35,489	1.9

solely by noise. The procedure of adding noise to the structure factors was described in Sect. 3.1, the noise control parameter was set to $p_1 = 0.222$, because this noise level generates approximately as many gross residual electrons (7.36) as exist in the experimental data after the MM (7.21). This provides a check of how the RDA-plot would appear in the case when all errors come from noise only at a constant experimental R-value (Tables 2 and 3).

The maximum values in h, k, and l were 21, 21, and 24, which leads to $n_x = 42$, $n_y = 42$, $n_z = 48$ for the residual density grid. The expectation value $\langle d^f(0) \rangle$ from Equation (14) is calculated to be 2.83198. This leads to percentage features of 5.3% (IAM), 4.4% (MM, exp.), and 1.9% (MM, theo.). For RDA-plots, see Fig. 6.

The MM reduced the features by only approximately 0.9%. This is in accordance with the small reduction in e_{gross} from 11.55 e (IAM) to 7.21 e (MM) and with the final distribution of $d^f(0)$, which is not parabolic. The little progress is due to a small disorder of the whole molecule. The corresponding occupation factor was, however, smaller than 1% and could not be refined (see also Fig. 7 and text in Sect. 3.4, where the same experimental data was used). The POF of the ideal model, however, is with 1.9% also higher than naively expected. A similar behavior was found in the previous section for e_{gross}, which did not reduce to zero despite using ideal data with zero noise. The same effect might take place here. The ideal expectation value is also calculated under the assumption of 100% completeness of the data, which is only approximately given.

From the previous discussion, it can be seen that the POF-values indicate that a reduction is still feasible, at least in principle.

For a comparison with other measures of features, the entropy S of the modulus of the residual density distribution was calculated according to

$$S = \sum_{\rho_0} |\rho_0| \ln(|\rho_0|). \tag{25}$$

The summation goes over all integer positive and negative multiples of 0.01 eÅ^{-3}. The entropy is then a positive number, which should be as small as possible for the residual density distributed as uniformly as possible. *S* decreases

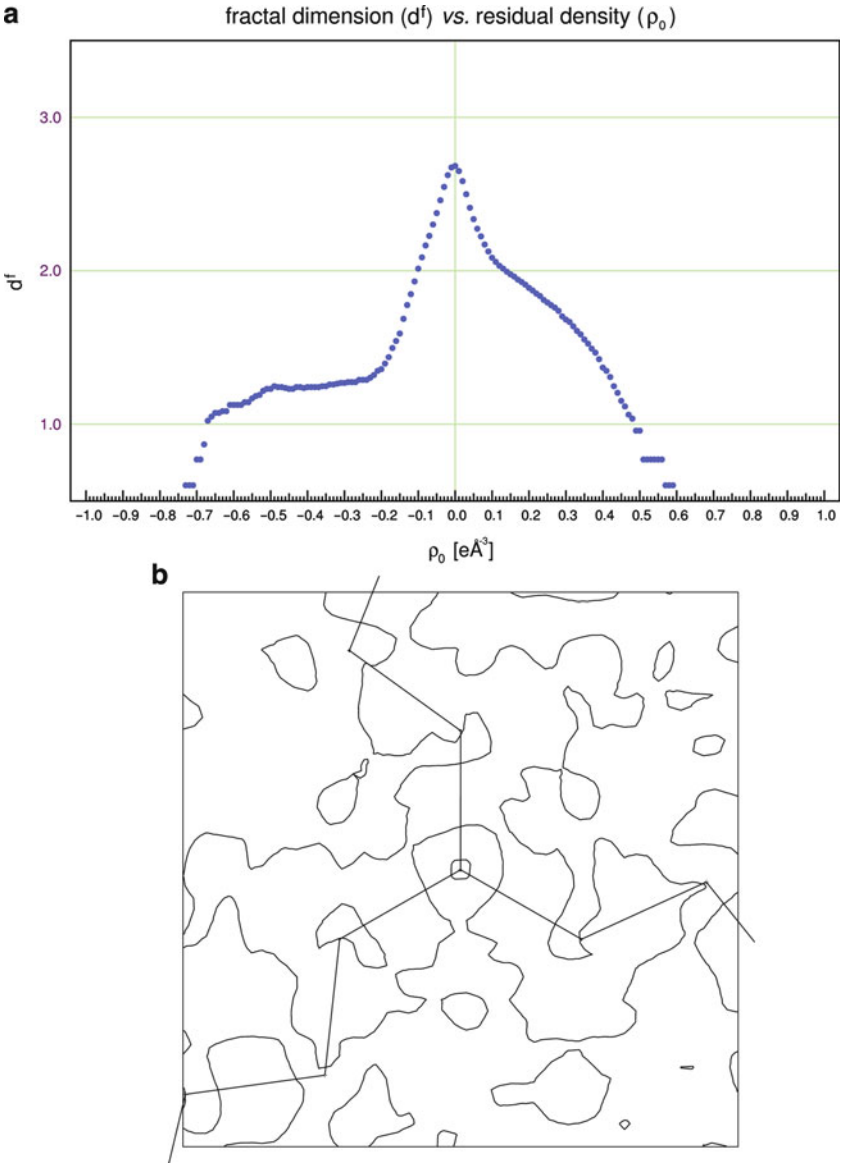


Fig. 6 RDA-plots (*left*) from a 42 × 42 × 48 grid and residual density contour representations of the zero residual density value (*right*) from a 50 × 50 × 1 grid from refinements against experimental data (*top*: IAM, *middle*: MM) and against ideal data (*bottom*). For simplicity, only the zero residual density line is depicted in the contour plots

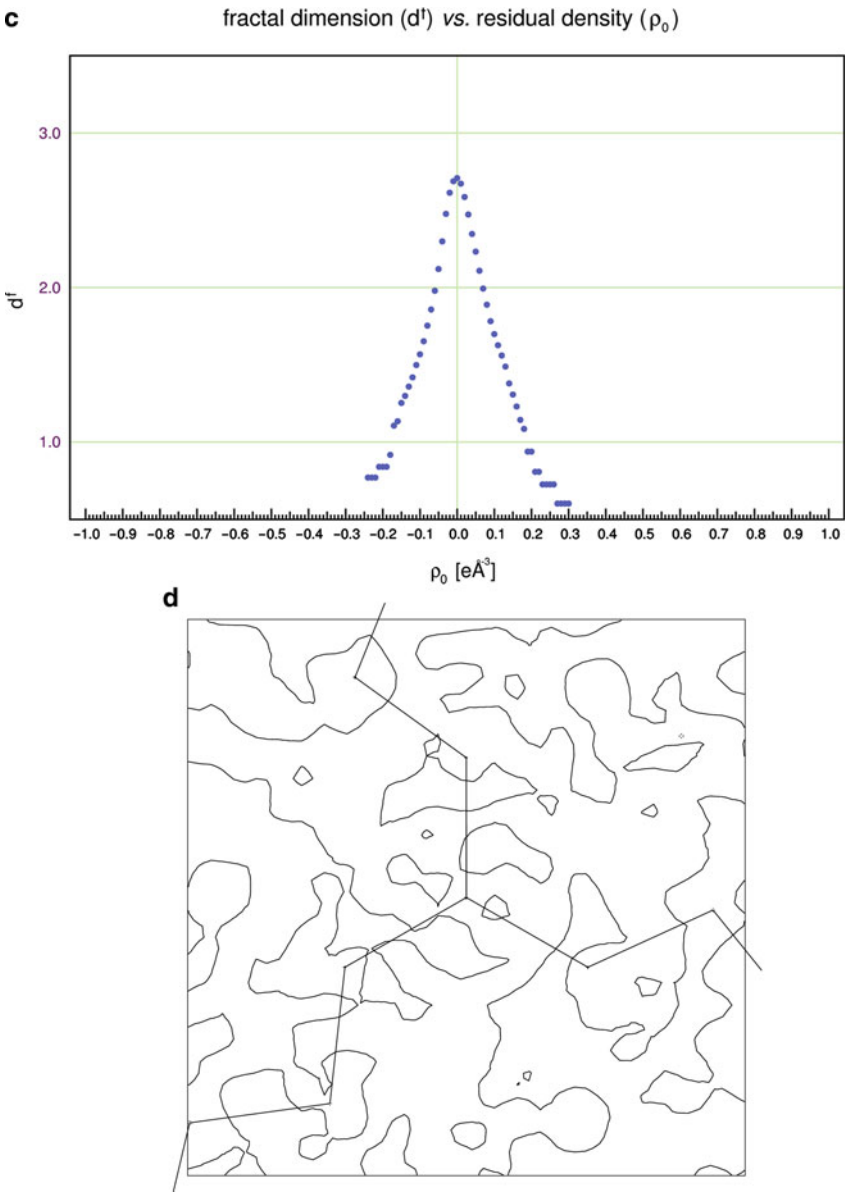


Fig. 6 (continued)

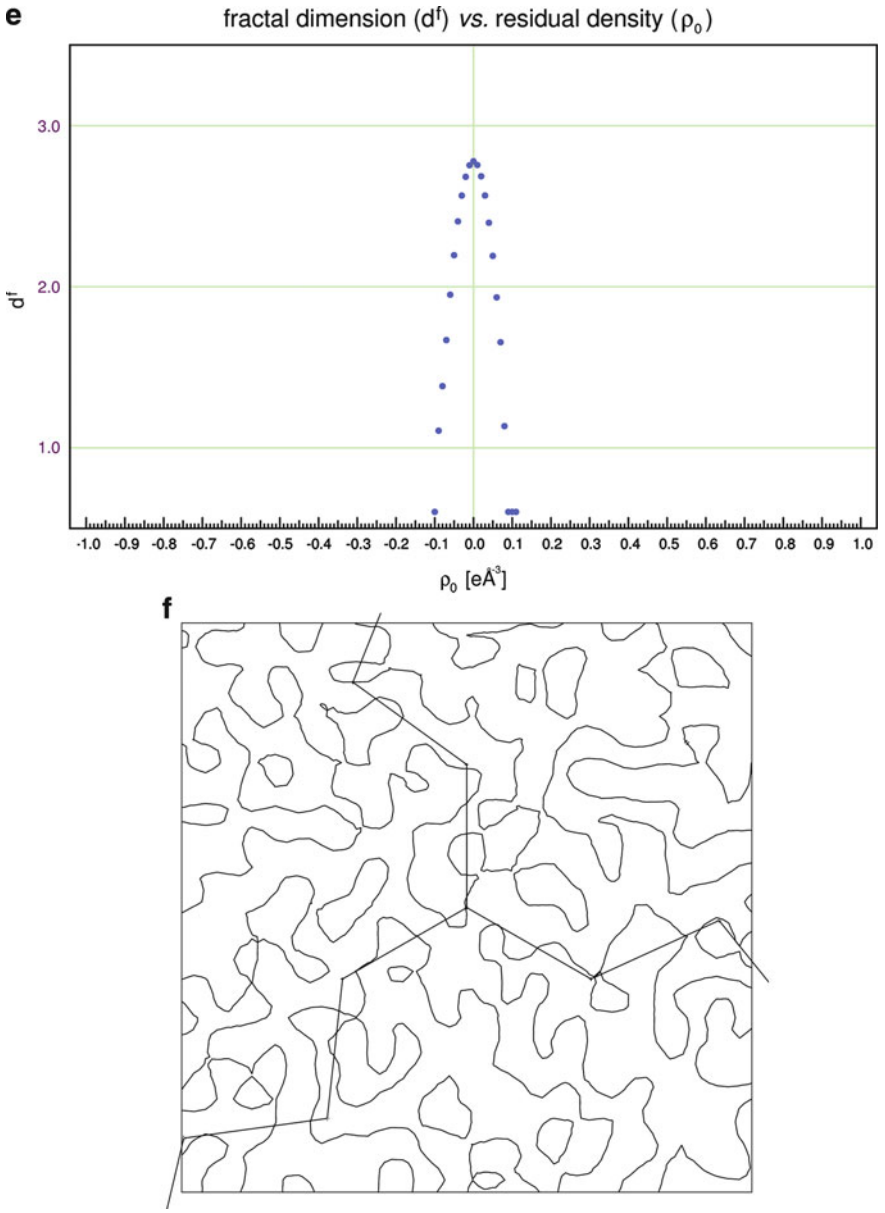


Fig. 6 (continued)

indeed from the IAM to the MM refinement from 69,564 to 30,574, indicating that the residual density distribution after the MM refinement is flatter than that after IAM refinement. In the MM refinement against theoretical data, however S increases again. Obviously, the definition of flatness from the RDA, which just uses $\Delta\rho_0$ and from an entropy functional S, which takes into account the whole distribution, have a different meaning: the p.d.f. corresponding to Fig. 6e is more peaked than the broad p.d.f. corresponding to Fig. 6c. This leads to the increase in the entropy for Fig. 6e. Please note that both have the same $wR_1|_{w=1}$ value as they have the same number of gross residual electrons. It must be concluded that the above functional prefers a non-Gaussian distribution of the residuals to minimize the entropy for a given, small, fixed total amount of errors.

3.4 . . . *to a Multipole Refinement*

To monitor the effect on the residual density distribution of a multipole model refinement against experimental data of $S(N^tBu)_3$ [8], a $50 \times 50 \times 50$ residual density grid of the whole unit cell was calculated with XDFOUR and analyzed after each individual refinement step with jnk2RDA. The distributions of the residual densities viewed through the glasses of the Residual Density Analysis approach are shown in Fig. 7 and the corresponding residual density descriptors in Table 4.

The scale factor was refined in each step. The multipole model started with the refinement of a scale factor and coordinates of all atoms, anisotropic displacement

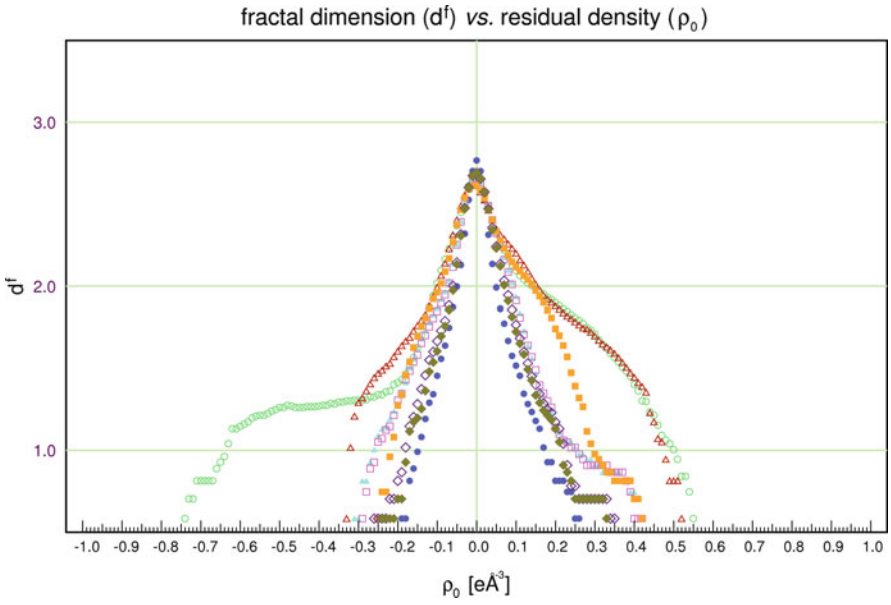


Fig. 7 Fractal dimension distribution for the individual steps of a multipole refinement on $S(N^tBu)_3$

Table 4 Residual density descriptors as obtained from an RDA of the individual steps of a multipole refinement on $S(N^tBu)_3$

	$d^f(0)$	e_{gross} [e]	$-\rho_{0,min}$ [eÅ^{-3}]	$\rho_{0,max}$ [eÅ^{-3}]	$\Delta\rho_0$ [eÅ^{-3}]
IAM	2.6681	11.54	0.74	0.55	1.29
$xyz + U_{ij}$ non-H ($\sin\theta/\lambda > 0.6$ Å^{-1})	2.7669	4.96	0.19	0.26	0.45
$P_v + \kappa$ non-H	2.6111	13.47	0.33	0.52	0.85
$P_v + P_{lm}$ all	2.6592	9.07	0.31	0.40	0.71
κ' non-H	2.6596	8.93	0.29	0.42	0.71
$xyz + U_{ij}$ non-H + RESET	2.6133	12.10	0.25	0.42	0.67
$P_v + \kappa + P_{lm}$	2.6918	7.38	0.26	0.35	0.61
$P_v + P_{lm} + \kappa + \kappa'$	2.6922	7.20	0.24	0.33	0.57

parameters for nonhydrogen and isotropic displacement parameters for hydrogen atoms from SHELX (*IAM*, green open circles) (Fig. 7). This refinement results in a large shoulder of the RDA-plot for negative residual density values due to the neglect of chemical interactions between the atoms. The error as given by e_{gross} equals 11.54 e. All following steps aim at reducing this error until noise is left only. The flatness (difference between maximum peak and deepest hole) $\Delta\rho_0 = 1.29$ eÅ^{-3} and the fractal dimension for the residual density value zero equals $d^f(\rho_0 = 0) = 2.6681$.

A high order refinement of the scale factor, the coordinates and the vibration parameters of the nonhydrogen atoms follows with $\sin\theta/\lambda > 0.6$ Å^{-1} ($xyz + U_{ij}$ non-H ($\sin\theta/\lambda > 0.6$ Å^{-1}), blue filled circles), which dramatically improves all residual density descriptors and the RDA-plot. This, however, is only due to data truncation and effective increase of the flexibility of the model by adjusting its parameters to fewer observations.

Thereafter the scale factor, monopole populations P_v and κ values for nonhydrogen atoms were refined against all data with $I > 3\sigma(I)$ for the whole resolution range ($P_v + \kappa$ non-H, red open triangles) leading to a pronounced reduction in the monopole values for the sulfur and carbon atoms, and to a slight increase of those for the nitrogen and hydrogen atoms. This step leads to a reduction of the negative residual density shoulder in the RDA-plot and it reduces $\Delta\rho_0$ to 0.85 eÅ^{-3}, however, e_{gross} increases to its maximum value (13.47 e) and $d^f(\rho_0 = 0)$ decreases to its minimum value in the whole refinement (2.6111), i.e., $\Delta\rho_0$ indicates a progress, whereas e_{gross} and $d^f(\rho_0 = 0)$ indicate a deterioration. These seemingly contradicting descriptors are interpreted as follows: the reduction in $\Delta\rho_0$ indicates that the new model that allows for charge transfer but not for atomic polarization is more appropriate than the spherical neutral atom model, whereas the increase in e_{gross} indicates the need for additional parameters to account for the asphericities. This is a counterexample for the often quoted sentence that anything can be fitted with an increasing number of model parameters, as in this case the *R*-values increase despite additional parameters.

In accordance with the expectations, all residual density descriptors indicate progress when also the multipoles are refined ($P_v + P_{lm}$ all, turquoise filled triangles) and the positive residual density shoulder in the RDA-plot decreases for the first time in the course of the refinement.

Only a slight change in the residual density distribution and its descriptors is induced by also allowing the aspherical deformation density expansion/contraction parameters κ' to change (κ' non-H, pink open squares).

In the next refinement step the hydrogen atoms were shifted to distances derived from neutron diffraction experiments (1.085 Å) and the scale factor as well as the coordinates of all heavy atoms and the anisotropic displacement parameters for the nonhydrogen atoms and isotropic displacement tensors for the hydrogen atoms were refined ($xyz + U_{ij}$ non-H + RESET, orange filled squares). The resulting flatness is $\Delta\rho_0 = 0.67$ eÅ^{-3}, the smallest value apart from the refinement where only a part of the experimental resolution was used, and which therefore is not comparable to the other refinements, but now the mono- and multipole values of the hydrogen atoms do not fit to the changed coordinates anymore causing a shoulder in the positive residual density region. This leads to an increase in e_{gross} toward 12.10 e (and a small value in $d^{\text{f}}(\rho_0 = 0) = 2.6133$).

When the scale factor, the multipole parameters for all atoms and the κ parameters for the nonhydrogen atoms are refined together, the model is given the opportunity to account for charge transfer and atomic density polarization for all atoms simultaneously ($P_{\text{v}} + \kappa + P_{\text{lm}}$, purple open rhombuses), which leads to a pronounced reduction in the total residual density as given by $e_{\text{gross}} = 7.38$ e, the smallest comparable value, and also $\Delta\rho_0 = 0.61$ eÅ^{-3} reaches its lowest comparable value.

Finally, the refinement of the scale factor, mono- and multipole parameters for all atoms together with a κ refinement for all and a κ' refinement for nonhydrogen atoms ($P_{\text{v}} + P_{\text{lm}} + \kappa + \kappa'$, olive filled rhombuses) leads to the largest value for $d^{\text{f}}(\rho_0 = 0) = 2.6922$, the lowest value for $e_{\text{gross}} = 7.20$ e, and the smallest value for $\Delta\rho_0 = 0.57$ eÅ^{-3}. From the RDA-plot, however, it can be seen that there is still potential for improving the model, as the final distribution still is not of parabolic shape. This may stem from a disorder of the whole molecule (rotation by 60° about the central axis through the sulfur atom) with an occupation factor smaller than 1%.

For a comparison with the more traditional residual density plots through the molecular plane, see Fig. 8, which shows the residual density before and after the multipole refinement. In the latter features remain in all three bisectors of the N–S–N angles. These indicate the positions of the N atoms in the disordered part.

3.5 ... to an Electron Density Reconstruction with the MEM

An independent alternative approach for the Multipole Model is the Maximum Entropy Method (MEM). In this approach, the dynamic density is reconstructed on a grid by maximization of an entropy functional subject to the constraint of fitting the diffraction data. Other additional constraints exist as well, such as the "prior derived F-constraints" [15] (PDCs), which are used for stabilizing the reconstruction process and leading to chemically and physically sensible electron density distributions.

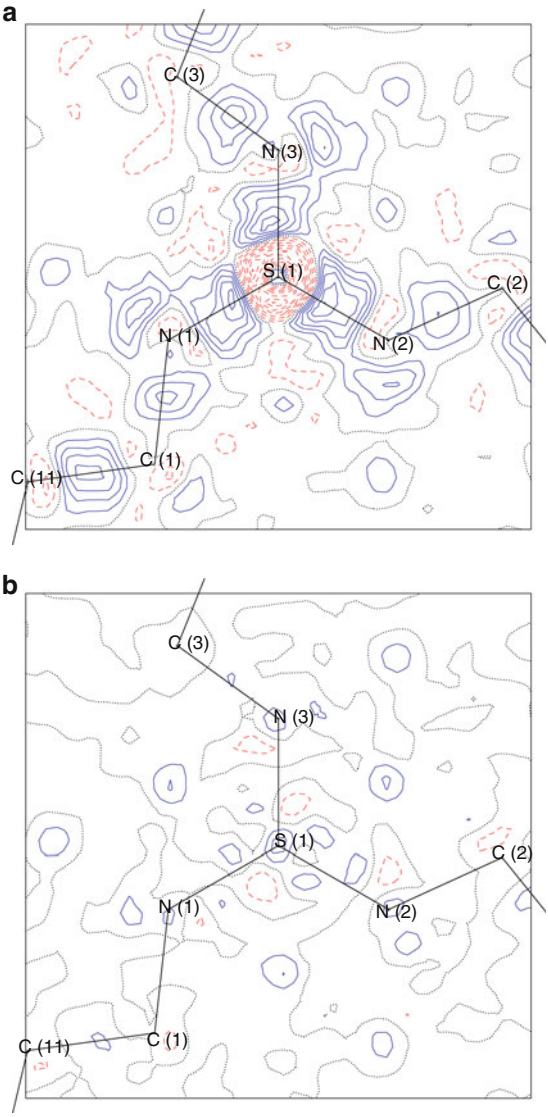


Fig. 8 Residual density representation as contour plots in the molecular plane. (**a**) IAM refinement. This model corresponds to the *first line* in Table 4. (**b**) After multipole refinement. This model corresponds to the *last line* in Table 4. *Blue solid lines*: positive residual density, *red dashed lines*: negative residual density, *gray dotted lines*: zero residual density. The contour spacing is 0.1 eÅ^{-3}

In the example given here, the resulting residual density grids from an application of the MEM to 23 K X-ray Mo-K_α data from α-glycine in the space group $P2_1/n$ with a resolution of $\sin\theta/\lambda = 1.15$ Å^{-1} [16] were subjected to an RDA. Residual density grids from a MEM reconstruction with PDC for the whole unit cell were analyzed.

To allow for a comparison of the experimental findings with idealized data, we used the coordinates and anisotropic displacement parameters from the experiment to generate a reflection file with the same resolution. Noise was added to the structure factors until the experimental number of gross residual electrons was approximately obtained. The reflection file was imported into XD and residual density grids were calculated from IAM refinements of these data. The RDA-plot of the experimental results with PDC (blue filled circles) and of the theoretical data (green open circles) is shown in Fig. 9.

Theoretical and experimental RDA-plots are in good agreement. Differences occur for $\Delta\rho_0$ and $d^f(0)$: both are smaller for the theoretical data and the difference in $d^f(0)$ is very large. In MM, it was frequently observed that $d^f(0)$ from the experimental data was smaller than from theoretical data. A possible explanation for these observations is that again the p.d.f. of the residual density is influenced by the entropy maximization procedure such that it leads to this spiked distribution. A similar behavior is observed for the entropy minimization of the modulus of the residual density distribution further above. That the resulting p.d.f. is more spiked in this case whereas it was broader in the above case need not be a contradiction as the preferred p.d.f. from the entropy functional may well be a function of the total error. Another possible explanation is that experimental noise was fitted by the MEM procedure, which would also lead to a redistribution of the residual density and possibly to a deviation from the Gaussian distribution. This example shows that an RDA may be useful for an evaluation of the properties of the entropy functional

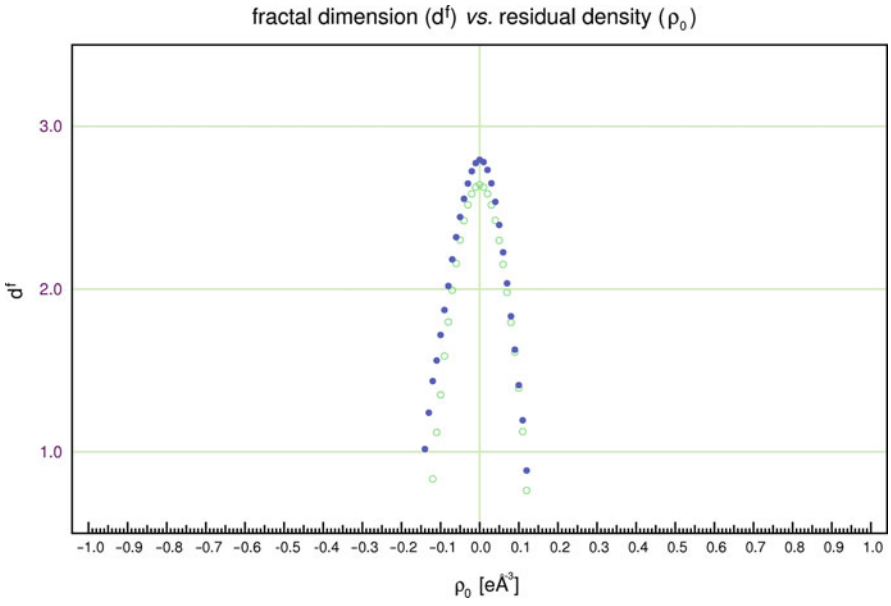


Fig. 9 RDA-plot of experimental MEM data (*blue filled circles*) and theoretical data (*green open circles*) with approximately the same number of gross residual electrons (3.4 e)

Table 5 Residual density descriptors for MEM data and theoretical data

	$d^f(0)$	e_{gross} [e]	$-\rho_{0,min}$ [eÅ^{-3}]	$\rho_{0,max}$ [eÅ^{-3}]	$\Delta\rho_0$ [eÅ^{-3}]
Exp. MEM data	2.7941	3.40	0.14	0.12	0.26
Theoretical data	2.5845	3.35	0.11	0.11	0.22

and for establishing appropriate stopping criteria for individual MEM processes (Table 5).

3.6 ... to Extinction Problems

Extinction is taken into account in SHELXL [17] by an empirical correction according to

$$F_{calc,corr} = k \frac{F_{calc}}{\left(1 + 0.001 \times \chi \times F_{calc}^2 \times \lambda^3 / \sin 2\theta\right)^{\frac{1}{4}}}, \tag{26}$$

with the scale factor k, the wave length λ, the diffraction angle θ, and the extinction coefficient χ. For $\chi = 0$, no extinction correction occurs.

To study the effects of extinction, the experimental data ($\lambda = 1.54178$ Å, 100 K) of bullvalene trisepoxide ($C_{10}H_{10}O_3$) [18] were treated in two different ways: in one IAM refinement, the data were not corrected for extinction effects and in the other case the data were corrected for extinction effects. The resulting reflections were used in XD [3] for least-squares refinements against F^2. The model used was again a conventional, spherical atom model with anisotropic harmonic thermal motion parameters, and exclusion of anomalous dispersion and extinction refinement (MODEL −2 2 1 0). Residual density grids with 50 × 50 × 50 points were calculated with XDFOUR and analyzed with jnk2RDA.

Figure 10 shows the distributions in terms of $d^f(\rho_0)$ (RDA-plots) and Table 6 gives the residual density descriptors.

The effect of the extinction correction ($\chi = 0.04935(240)$) is large. The difference between the largest peak and deepest hole $\Delta\rho_0$ (flatness) decreases from 0.65 to 0.28 eÅ^{-3}, while the total error as given by e_{gross} decreases from 30.26 to 11.03 e. Please note that this change of almost 19 gross residual electrons due to modeling of extinction is much larger than the total change in gross residual electrons induced by a multipole refinement of in total 4.3 e as discussed for $S(N^tBu)_3$ in Sect. 3.4. As a consequence, the extinction needs to be modeled very accurately to avoid systematic errors. It is again the very special needs in charge-density studies, which shine through this discussion. It may well be the case that an extinction correction according to Equation (26) is absolute satisfactory for standard structure determinations, but this does not imply that it is necessarily satisfactory for charge-density studies, too.

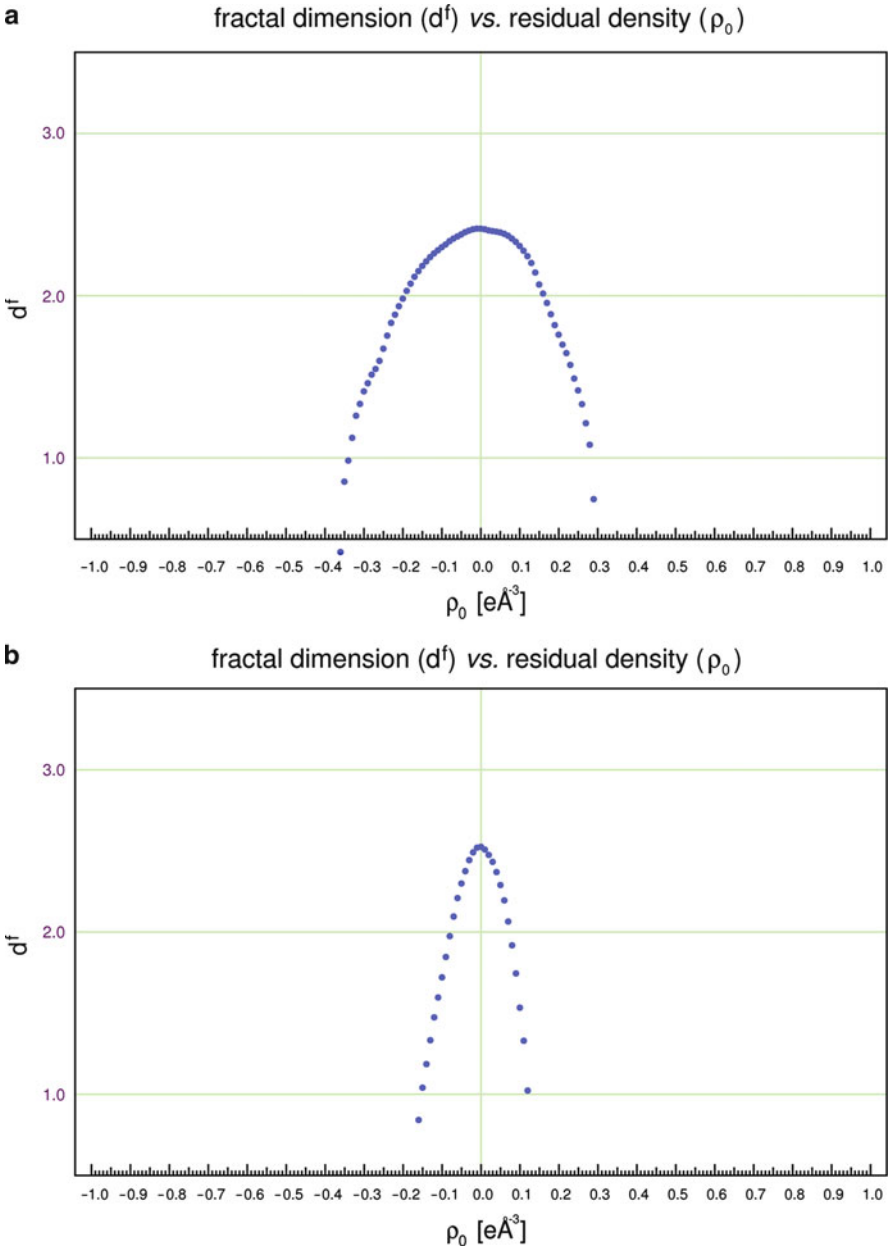


Fig. 10 Fractal dimension distribution of the residual density of bullvalene trisepoxide after IAM refinement without extinction correction (**a**), and with empirical extinction correction (**b**)

Table 6 Residual density descriptors applied to experimental data for bullvalene trisepoxide to study the effects of the extinction correction as implemented in SHELXL

Correction for extinction	$d^f(0)$	e_{gross} [e]	$-\rho_{0,min}$ [eÅ^{-3}]	$\rho_{0,max}$ [eÅ^{-3}]	$\Delta\rho_0$ [eÅ^{-3}]
Off	2.4129	30.26	0.36	0.29	0.65
On	2.5228	11.03	0.16	0.12	0.28

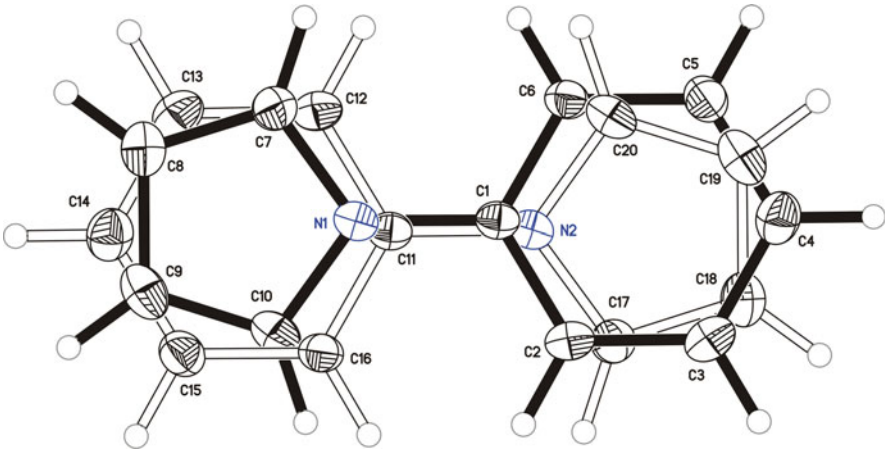


Fig. 11 *N*-phenylpyrrole (PP) in its disordered state. The main domain with occupation factor 0.9 is plotted in *black*, the minor domain in *white*. The anisotropic displacement parameters are shown at the 50% probability level

3.7 . . . to Statistical Disorder

Disorder of solvent, molecules, or parts of molecules is of particular importance in charge-density studies. As has been mentioned earlier, the presence of unmodeled disorder influences the model parameters and in this way distorts the density model in a way not easy to control. Conversely, if disorder is taken into account and if it can be shown that the remaining residual density does not contain any structural information except for noise (at the given experimental resolution), one has a stopping criterion for the refinement and a quality control. Of course, this still does not prove the model to be correct, but the converse is true: features in the residual density disprove the model to be correct. Therefore, among all models only those with a statistical residual density are acceptable.

The example considered here is from 100 K Mo-K_α data of *N*-phenylpyrrole $H_4C_4N(C_6H_6)$ crystallizing in the noncentrosymmetric orthorhombic space group $P2_12_12_1$ with one molecule in the asymmetric unit, which is disordered by a 2-fold rotation axis (see Fig. 11). The occupation factors are approximately 0.9 and 0.1. More details can be found in [1, 19, 20].

A conventional multipole refinement of the dominant domain and a multipole refinement taking the disorder into account have been performed and residual density grids have been calculated and analyzed.

A 100 × 60 grid was used for the plane with edge lengths 10 and 6 Å in Fig. 12a and a 56 × 76 × 176 grid was used for the RDA-plot in Fig. 12b. Both graphics clearly indicate that the residual density is far from being flat and featureless.

The unmodeled disorder is clearly visible as positive residual density between the atomic positions of the model and as negative residual density (red dashed lines) around the atomic positions. The positive residual density at the atomic positions of the first domain stems from too large anisotropic displacement parameters, which artificially increased during the refinement to account for the other (unmodeled) domain. Too large values for the U_{ij} generate circular areas of negative residual density around the atomic positions. These are also seen in Fig. 12a. A study on how unphysical values of anisotropic displacement parameters affect the whole residual density distribution and planes through the molecule can be found in [1]. The unmodeled domain of PP contributes to the positive residual density shoulder in Fig. 12b, whereas the too large occupation factor (1.0 instead of 0.9) contributes most to the negative residual density shoulder.

After taking the statistical disorder into account, the residual density distribution changes drastically. In Fig. 13, again a plane through the molecule is depicted as well as the RDA-plot taking into account the residual density in the whole unit cell. The fractal dimension $d^f(\rho_0 = 0)$ increases from 2.5770 to 2.6286, which, in view of the logarithmic scale involved in the definition of the fractal dimension of the residual density, is interpreted as a rather large change. The total error decreases from 16.08 to 6.74 gross residual electrons. Finally, the difference between maximum peak and deepest hole decreases from 0.98 to 0.27 $e\text{Å}^{-3}$. The additional refinement of the second domain describes the total electron density in the whole unit cell to a very good degree. A good description, i.e., a description that fits the experimental observations, is a necessary condition for an acceptable model. Whether or not this model is physically acceptable (and chemically meaningful) is out of the scope of the residual density analysis, for questions of this kind clearly physical, and not only statistical, measures of quality need to be developed and applied and the resulting density is to be analyzed with respect to information about electrostatic forces (Tables 7 and 8).

3.8 . . . to an Anharmonic Motion Problem

One main application of the RDA(-plot) is to prove absence of features in the residual density distribution to show that the density and thermal motion model is among the acceptable models. This was also the way the RDA was applied in the preceding section about static statistical molecular disorder.

Here, we describe in detail a case where the thermal nuclear motion model was crucial. The experimental results are already discussed in a chapter of Electron

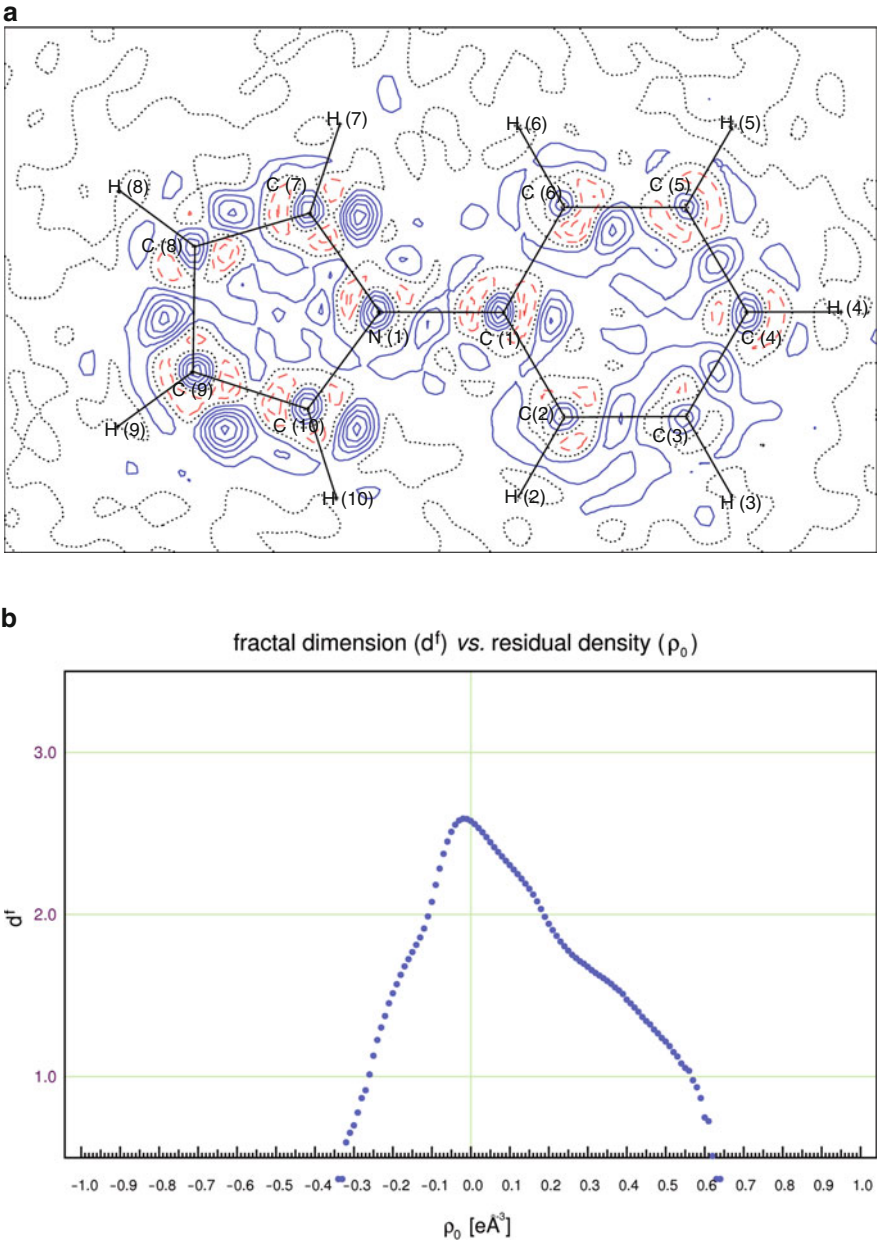


Fig. 12 Residual density of PP after refinement of the main domain. (**a**) *blue solid lines*: positive residual density, *red dashed lines*: negative residual density, *gray dotted lines*: zero residual density. Contour spacing 0.1 eÅ^{-3}. (**b**) RDA-plot. The values of the residual density descriptors are given in Table 7

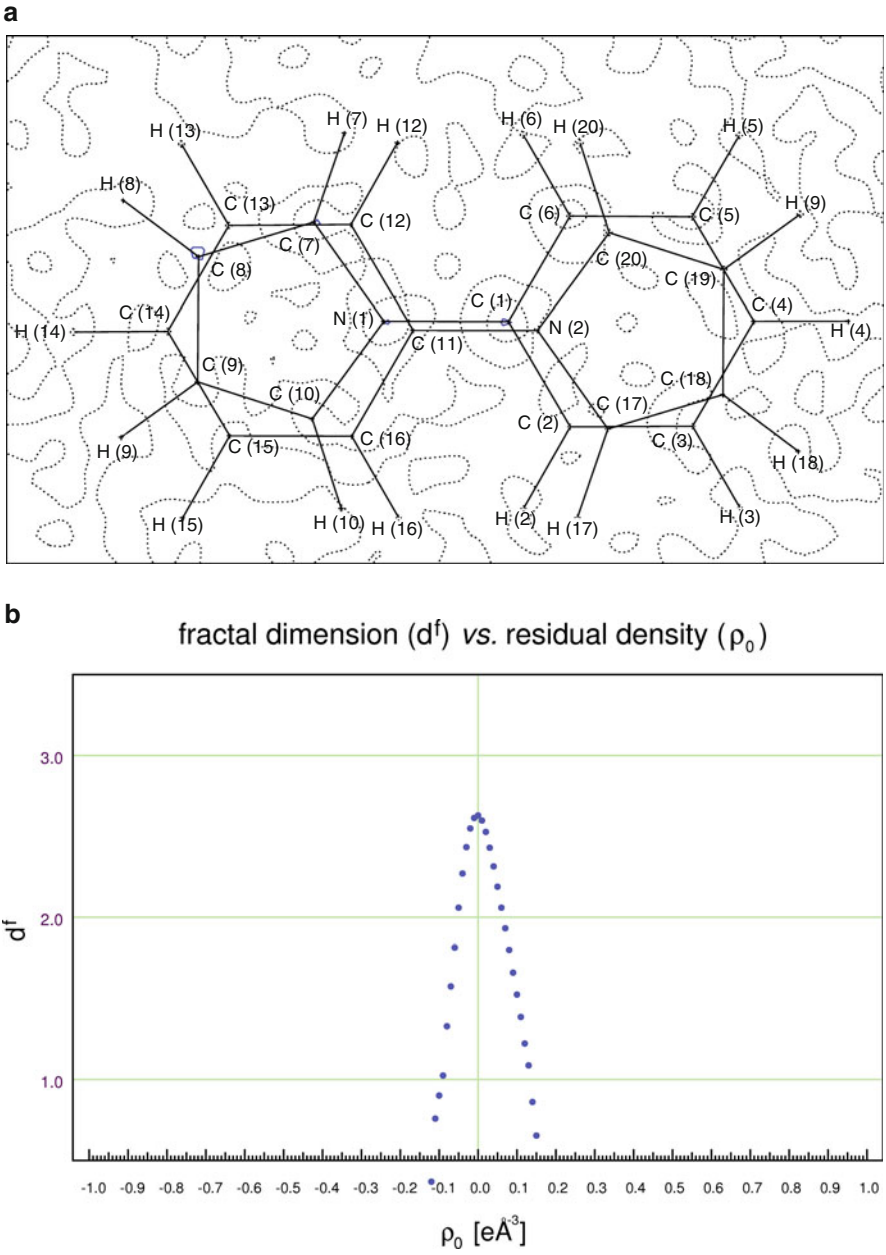


Fig. 13 Residual density of PP after refinement of main and minor domain. (**a**) *blue solid lines*: positive residual density, *gray dotted lines*: zero residual density. Contour spacing 0.1 eÅ^{-3}. (**b**) RDA-plot. The values of the residual density descriptors are given in Table 7

Table 7 Residual density descriptors for *N*-phenylpyrrole with inclusion and exclusion of minor domain refinement

Refinement of disorder	$d^f(0)$	e_{gross} [e]	$-\rho_{0,min}$ [eÅ^{-3}]	$\rho_{0,max}$ [eÅ^{-3}]	$\Delta\rho_0$ [eÅ^{-3}]
No	2.5770	16.08	0.34	0.64	0.98
Yes	2.6286	6.74	0.12	0.15	0.27

Table 8 Residual density descriptors, crystallographic *R*-value, and parameter *R*-values for a refinement of a multipole model excluding (first row) and including (second row) anharmonic nuclear motion at the P and Al atom, respectively

3rd ord. GC	4th ord. GC	R_1 [%]	POF [%]	$d^f(0)$	$\Delta\rho_0$ [eÅ^{-3}]	e_{gross} [e]	R_{dip}	R_{quad}	R_{octu}	R_{hexa}	R_{sum}
Off	Off	1.56	1.34	2.68	0.54	16.00	0.07	0.06	0.10	0.11	0.30
P/Al	P	1.54	1.29	2.68	0.27	15.83	0.03	0.03	0.07	0.12	0.19

Density and Chemical Bonding I. The chemical interpretation of the resulting electron density is also discussed there.

Here, we focus on the methodological aspect. The question to be tackled is: provided, anharmonic motion is definitely present and provided the electron density can be modeled appropriately by a multipole expansion and provided anharmonic nuclear motion is neglected in the model: are there signs in the refinement warning us for this deficiency? The application of such knowledge is straightforward.

To answer these questions, theoretical structure factors from a known multipole model are used and the density is convoluted with anharmonic nuclear motion parameters for the two heaviest atoms P and Al as determined from the experiment and the corresponding anisotropic displacement parameters for the remaining atoms. Gaussian noise is added to the structure factors. The amount of noise added is such that the experimental *R*-value is fitted. The data correspond to a resolution of $\sin\theta/\lambda = 1.15$ Å^{-1} and the experimental temperature was 100 K. More information about the experimental setup and conditions leading to this model can also be found in the above-mentioned chapter and in [21], more about the theoretical approach and all GC coefficients with estimated standard uncertainties can be found in [22]. As the true density- and thermal motion parameter values were known due to using simulated data, we were also able to monitor the difference between the true density- and thermal motion parameters and those derived from a least-squares refinement of a reduced model that lacks anharmonic motion parameters to different degrees, e.g., lack of 4th order at the P atom, lack of 4th order at the P, and of 3rd order at the Al atom and total neglect of anharmonic motion. This monitoring is done by the parameter *R*-value, which is defined as the sum of absolute difference between the true reference model density parameters and those resulting from a least-squares fit. As we are mainly interested in the aspheric distribution around the nuclei and as the parameter *R*-value is dominated by little changes in the monopole parameter, we excluded the monopole from the calculation of the parameter *R*-value. Therefore, a parameter *R*-value of zero for an

atom would mean that this atom has the same density shifts as the reference atom, but it still may have a different monopole value. The parameter R-value is defined in the following equation, where "dip" abbreviates dipole, "quad" quadrupole, "octu" octupole, and "hexa" hexadecapole.

$$\begin{aligned} R_{\text{para}} &= R_{\text{dip}} + R_{\text{quad}} + R_{\text{octu}} + R_{\text{hexa}}, \\ R_{\text{lm}} &= \sum_{m=-l}^{+l} \left| P_{\text{lm,reference}} - P_{\text{lm,model}} \right|, \quad l = 1,\ 2,\ 3,\ 4. \end{aligned} \tag{27}$$

When a multipole refinement is performed, one usually tries to avoid anharmonic nuclear motion refinement, as this might introduce severe parameter correlations. Figure 14 shows the deepest hole and highest peak in the residual density together with the resulting VSCCs close to the P atom from such a refinement excluding anharmonic motion.

The residual density peak and hole are closer to the P atom than the VSCCs and they are lying on a line through the nucleus of the P atom. Due to the close vicinity of the residual density features and the VSCCs naturally the question arises whether these are influenced by the peak and hole. We will see that this is indeed the case: two of the VSCCs are artificial and occur only due to the neglect of present anharmonic nuclear motion of the P atom. When anharmonic motion is taken into account and the Gram–Charlier parameters are refined, these artificial VSCCs disappear and the absolute values of residual density peak and hole become smaller.

Figure 15 compares the final model that takes anharmonic motion into account and the model that neglects anharmonic motion.

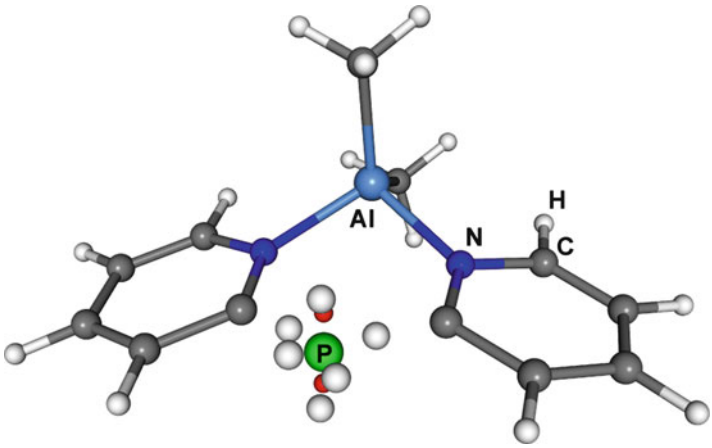


Fig. 14 Largest peak (0.23 eÅ^{-3}) and deepest hole (−0.21 eÅ^{-3}), in the residual density (*small red spheres*), and VSCCs (*white spheres*) in the vicinity of the P atom (*green sphere*) after a multipole refinement excluding anharmonic motion

The neglect of 3rd and 4th order Gram–Charlier coefficients at the P atom and of 3rd order at the Al atom has only a little effect on the *R*-values: the reference model R_1-value was 1.54% and the R_1-value corresponding to the refinement neglecting 3rd order Gram–Charlier coefficients was 1.56%, the total difference being only 0.02%. For the introduction of 35 new parameters (10 for each set of 3rd order coefficients and 15 for the 4th order), this is a rather small gain in the R_1-value.

The R_1-value, however, is not the only quality criterion to be applied in charge-density studies. There is a distinct effect on the residual density distribution as given by the RDA-plots: neglect of Gram–Charlier coefficients leads to shoulders in the fractal dimension distribution of the residual density (see Fig. 15e, f).

A more detailed analysis shows that the neglect of 4th order GC coefficients at the P atom induces one artificial VSCC and the neglect of 3rd order GC coefficients induces another one. The neglect of 3rd order GC coefficients is predominantly causing the residual density peak and hole at the P atom, which also causes the "shashlik"-like residual density iso-surface with alternating positive and negative values and with the P atom in a nodal plane.

The small difference in the crystallographic R_1-values between the models is reflected in small or (to the given figures) no differences for $d^f(0)$ and e_{gross}.

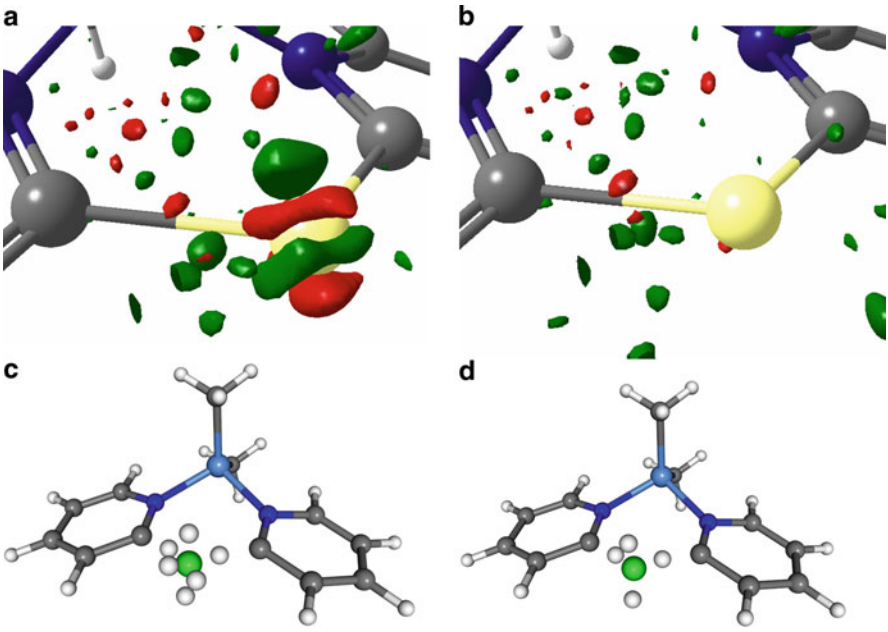


Fig. 15 Comparison of refinements against theoretical data neglecting anharmonic motion (*left*) and taking anharmonic motion into account (*right*). (**a**) and (**b**): residual iso-density representation at the 0.09 (*green*) and the −0.11 (*red*) eÅ^{-3} residual density level. (**c**) and (**d**): Valence shell charge concentrations (*white spheres*) in the vicinity of the P atom. (**e**) and (**f**): RDA-plots

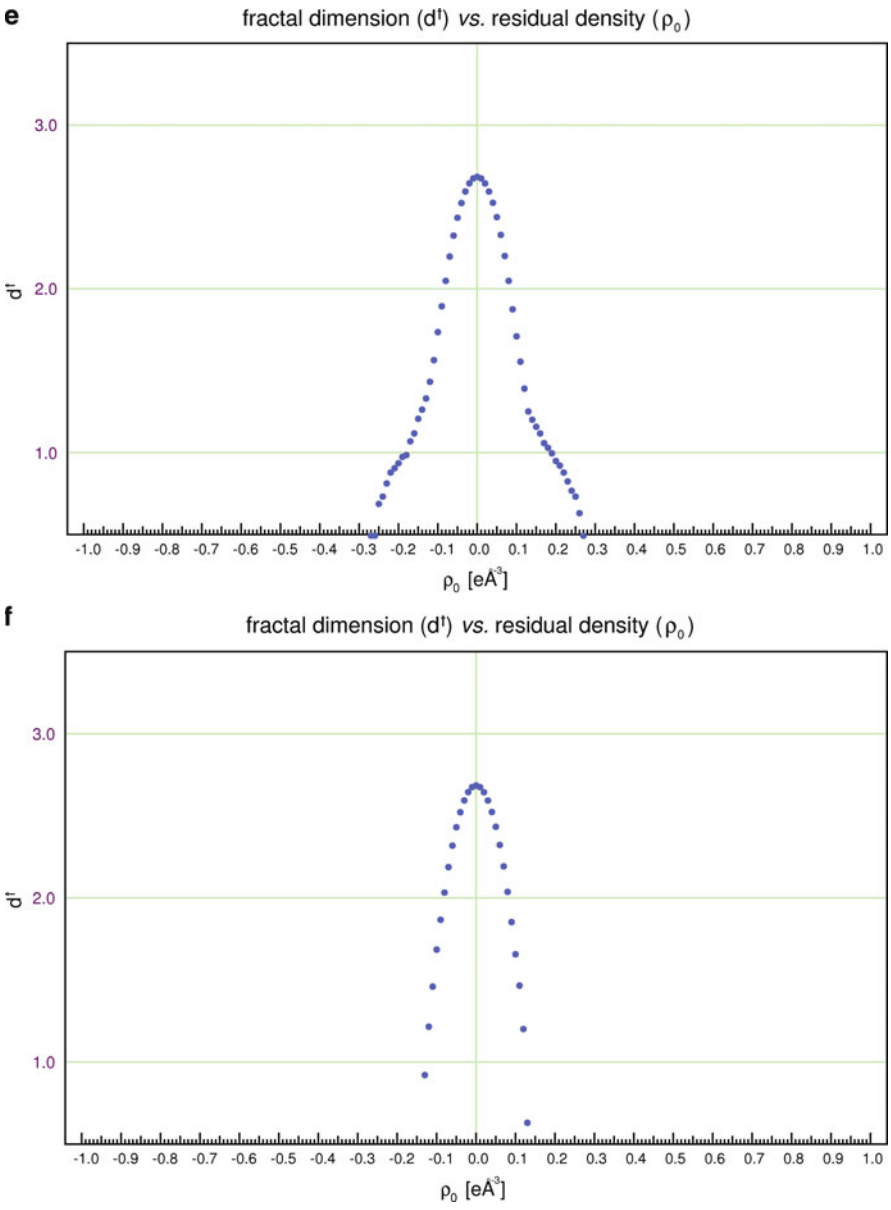


Fig. 15 (continued)

The parameter R-values differ more and for the model excluding anharmonic nuclear motion all individual parameter R-values but R_{hexa} are larger compared to the full model. As the parameter R-values for the appropriate model (second row) are solely determined by the noise, this may be a sign that the hexadecapoles fit experimental noise efficiently in this case. May this be as it is, the total parameter R-value is larger for the model lacking anharmonic nuclear motion. Introduction of anharmonic nuclear motion parameters reduces $\Delta\rho_0$ from 0.54 to 0.27 eÅ^{-3}.

The neglect of anharmonic nuclear motion may lead to artificial VSCCs in the electron density. When the highest residual density peak and deepest hole are on a straight line through a nucleus, this may be an indicator for the neglect of 3rd order GC coefficients. A shashlik-like pattern in the residual density with alternating signs and with an atom in a nodal plane may also indicate neglect of 3rd order GC coefficients.

4 Future Developments

The RDA is only a little step toward more transparency in the detection, classification, and elimination of density model and thermal motion model inadequacies, in data processing and data reduction inadequacies, and so forth. As a comparative method, it cannot state that either the data or the model is wrong, but it can detect systematic discrepancies between data and model. For the plausibility of the resulting density and thermal motion models, other than purely statistical measures of quality need to be used.

4.1 Physical Measures of Quality

The statistical measures only give a degree of agreement between the theoretically and experimentally obtained values. Nothing is said about the physics. Therefore, a model with very high agreement still could be "unphysical."

For structure determination, the statistical measures are completely sufficient; however, when describing the electron density distribution in great detail, it might be helpful to have additional measures of quality and/or additional knowledge, which could be used as extra observations in the refinement. Why not using physical measures of quality?

For example, due to the thermal motion (and the Heisenberg uncertainty principle), the nuclei in the crystal change their place all the time slightly, however, on macroscopic time scales they do not move (crystal damage is not considered here). The forces acting on the atoms are such that they drive the atoms back to their mean locations, which do not change with time. It is an equilibrium situation. An equilibrium situation is established when the macroscopic properties do not change with the course of time. Microscopically, however, changes may occur all the time.

Now, if the density and thermal motion models result in strong net forces acting onto the nuclei, these forces would shift the nuclei until equilibrium was obtained. The model is unphysical in this case, as the crystal is in equilibrium.

Of course, it would be desirable to have vanishing electrostatic forces as the *result* of the refinement without forcing the model to obey such constraints in the first place. It must be said, however, that the experimental refinement procedure is not of that quality and virtually every experimentally obtained model is prone to violate equilibrium conditions.

On the other hand, obeying the equilibrium conditions still does not *prove* the model to be the best possible model. But, and that is important, the violation of any equilibrium condition proves the model to be unphysical and therefore unacceptable for describing an equilibrium system. That is why physical measures of quality should be available in the future.

4.2 Further Tools for Analysis, Advanced Theories, and Adjustments to Charge-Density Needs

There has been a large progress in X-ray detection with the area detectors and in the data processing capabilities by the improved IT technologies. These have been used to make data collection and standard structure determination faster, and to enable charge-density studies, however, for the special needs of charge-density applications, there needs to be a similar progress in the data processing, error analysis, and model building. Model building again comprises not only density and thermal motion models but also the scattering model, which at present neglects TDS and Renninger reflections. In future, it might be necessary to identify and quantify the Renninger contributions to individual reflections and to handle the TDS more accurately if we want to be in control of the sources of error. It might also be necessary to investigate and recalibrate incident beam corrections and photon conversion rates for individual experimental setups. Whether this is necessary or not depends on how accurate and precise the experimental results are required to be. Also, scaling and merging techniques should be revised to assure appropriateness for charge-density studies. One example was given above where it has been shown that an empirical extinction correction influences the data such that errors of 10% in this empirical correction would amount to approximately 4 gross residual electrons, which may be the whole difference between an IAM and a multipole model refinement.

Another example is given here: Suppose a weighting scheme is used for identifying and subsequent downweighting of outliers. What does this imply for weak and zero intensity observations? As the minimum error is given by the counting statistics, this obviously means that there is no easy way of identifying outliers for weak and zero intensity observations, as these do not have a sharp expectation value. The concept of an outlier cannot be easily applied to weak and zero intensity observations. This holds also in the case of zero background. If outlier treatment is

applied nevertheless, there is a danger of obtaining too significant reflections by an artificial reduction of the standard deviations.

Another example: in [23], a scaling scheme is given by iterative least-squares fitting to minimize

$$\chi^2 = \sum_{\mathbf{h}} w_{\mathbf{h}} \left(\frac{|F_{\mathbf{h}}|_{\text{meas}}}{\left\langle |F_{\mathbf{h}}|^2 \right\rangle^{\frac{1}{2}}} - k^{-1} \left\langle \mathrm{e}^{-2\pi^2 \langle u^2 \rangle / d_{\mathbf{h}}{}^2} \right\rangle \right)^2, \tag{28}$$

where the denominator in the first term in brackets is evaluated by an application of Wilson's statistics [24]. The scaling factor to be determined is k and the displacement parameters are hidden in the exponent. When in a charge-density experiment, the data has been scaled according to Equation (28) and there is no disorder and a high quality of the model, then one could in a postprocessing step rescale the data where Wilson's statistic is replaced by the statistics derived from the model. This would lead to a correction and to a self-consistent procedure. The reader might think that the scaling is then biased toward the model used and therefore the proposed procedure should not be accepted. One can see it the other way round, too: If scaling is performed according to (28), then the scaling is biased toward Wilson's statistics, which implicitly assumes uniform random atoms at rest on no special positions. Wilson's statistics might be a good starting point, when not much is known about the structure. However, when the structure is known in great detail, Wilson's statistics might not be a good assumption any more, in particular when atomic numbers vary heavily or when heavy atoms are on special positions. Then the known IAM structure is a better prior knowledge than Wilson's statistics. In any case, the proposed procedure is applicable only as a postprocessing step for very high quality data and very good density models in the absence of disorder.

The discussion of this section shows that there is a shift from the straightforward application of (over?)-simplified scattering theories toward complex simulations and deeper analysis of potential and actual sources of error. This effort might not be necessary for every charge-density experiment. On the other hand, if it allows the determination of the electron density more accurately also for weakly diffracting compounds, it might well be worth the effort. As a by-product, there is much to learn about the physics of the scattering process, machines and detectors, about the statistics useful in data reduction processes and so forth.

4.3 A Common Effort

The most important question is that of learning together. As an individual one just adds to the work, but being part of a larger community acting in concert means that the individual efforts will be multiplied and the rewards will be for the individuals contributing and for the charge-density community as a whole. This is the spirit of

the RDA. And this is the reason we ask anyone who applies the RDA, to return at a time a nice example of a problem that was described or solved with it. These examples will be collected and published as soon as enough interesting problems (and solutions!) are collected. The focus is mainly on the characteristic distribution of $d^f(\rho_0)$, but any example is welcome. This will help beginners to become more experienced and it will help the experienced to identify and possibly remove, but at least to discuss sources of systematic errors, leading to more reliable charge-density distributions for all.

Acknowledgments Thanks to Dr. Leusser for providing the refinement strategy for $S(N^tBu)_3$. Thanks to Dr. Herbst-Irmer for fruitful discussions on anharmonic motion. This work was supported by the DFG (grant HE 4573/3-1).

References

1. Meindl K (2008) Residual density validation and the structure of Labyrinthopeptin A2. Georg-August-Universität, Göttingen
2. Henn J, Meindl K (2010) Acta Crystallogr A66:676–684
3. Volkov A, Macchi P, Farrugia LJ, Gatti C, Mallinson PR, Richter T, Koritsanszky T (2006) XD2006 – A computer program package for multipole refinement, topological analysis of charge densities and evaluation of intermolecular energies from experimental and theoretical structure factors
4. Dittrich B, Koritsánszky T, Luger P (2004) Angew Chem Int Ed 43:2718–2721
5. Renninger M (1937) Z Physik 106:141–176
6. Massa W (2007) Kristallstrukturbestimmung. Teubner, Stuttgart
7. Mandelbrot BB (1982) The fractal geometry of nature. W. H. Freeman, New York
8. Leusser D, Henn J, Kocher N, Engels B, Stalke D (2004) J Am Chem Soc 126:1781–1793
9. Shannon CE (1949) Proc IRE 37:10–21
10. Meindl K, Henn J (2008) Acta Crystallogr A64:404–418
11. Jelsch C, Guillot B, Lagoutte A, Lecomte C (2005) J Appl Crystallogr 38:38–54
12. Guillot B, Viry L, Guillot R, Lecomte C, Jelsch C (2001) J Appl Crystallogr 34:214–223
13. Jayatilaka D, Grimwood DJ (2003) Lect Notes Comput Sci 2660:142–151
14. van Smaalen S, Palatinus L, Schneider M (2003) Acta Crystallogr A59:459–469
15. Palatinus L, van Smaalen S (2005) Acta Crystallogr A61:363–372
15. Netzel J, Hofmann A, van Smaalen S (2008) Cryst Eng Commun 10:335–343
16. Sheldrick GM (2008) Acta Crystallogr A64:112–122
18. Liang S, Lee C-H, Kozhushkov SI, Yufit DS, Howard JAK, Meindl K, Rühl S, Yamamoto C, Okamoto Y, Schreiner PR, Rinderspacher BC, de Meijere A (2005) Chem Eur J 11:2012–2018
19. Meindl K, Henn J, Kocher N, Leusser D, Zachariasse KA, Sheldrick GM, Koritsanszky T, Stalke D (2009) J Phys Chem A 113:9684–9691
20. Kocher N (2003) Experimental charge density studies of highly polar bonds. Bayerische Julius-Maximilians-Universität, Würzburg
21. Henn J, Meindl K, Oechsner A, Schwab G, Koritsanszky T, Stalke D (2010) Angew Chem Int Ed 49:2422–2426
22. Meindl K, Herbst-Irmer R, Henn J (2010) Acta Crystallogr A66:362–371
23. Blessing RH (1999) DREAR – Data reduction and error analysis routines for ab initio phasing and for accurate electron density mapping. IUCr Crystallographic Computing School "Frontiers in Computational Crystallography", Hinxton, Cambridge, England
24. Wilson AJC (1949) Acta Crystallogr 2:318–321

Struct Bond (2012) 147: 193–286
DOI: 10.1007/430_2010_31
© Springer-Verlag Berlin Heidelberg 2010
Published online: 3 December 2010

The Source Function Descriptor as a Tool to Extract Chemical Information from Theoretical and Experimental Electron Densities

Carlo Gatti

Abstract This chapter deals with the source function (SF) descriptor, originally put forth by Bader and Gatti back in 1998. After a brief review on how this descriptor is defined and what it physically represents, the various forms through which the SF may be analyzed are presented in some detail. The relationships between atomic SF contributions and chemical bond nature are analyzed in some prototypical cases, and the capability of the SF to neatly reveal π-electron conjugation directly from the electron distribution and independently from any MO scheme or decomposition is introduced. Applications of the SF to chemistry from the literature are reviewed and critically discussed, including the use of the SF to assess chemical transferability or to describe chemical bonding in challenging situations, like for instance the short-strong hydrogen bonds in π-conjugated frameworks or the metal–metal and metal–ligand interactions in the organometallic complexes. Comparison with the insight obtained from other bond topological descriptors is given, emphasizing the special role the SF has of being directly derivable from experimental electron density distributions and to so provide an ideal tool to compare experiment and theory. The robustness of the SF descriptor against changes in the models used to derive electron densities from theory of experiment is detailed. First results on using the SF to define an unambiguous full population analysis are outlined. The possible ways of further decomposing the atomic SF in chemically meaningful additive pieces, such as core and valence atomic contributions, are analyzed in view of their potential insight and degree of arbitrariness.

Keywords Chemical transferability · Electron conjugation · Local and nonlocal bonding descriptors · Metal–metal and metal–ligand bonds · Population analysis · Short-strong hydrogen bonds · Source function and chemical bonding · Theoretical and experimental electron densities

C. Gatti (✉)
Istituto di Scienze e Tecnologie Molecolari del CNR (CNR-ISTM) e Dipartimento di Chimica Fisica ed Elettrochimica, Università di Milano, Via Golgi 19, Milano 20133, Italy
and
Center for Materials Crystallography, Aarhus University, Langelandsgade 140, Aarhus C 8000, Denmark
e-mail: c.gatti@istm.cnr.it

Contents

1 Introduction 195
2 The Source Function: Basic Aspects 195
2.1 Mathematical Derivation and Physical Interpretations 196
2.2 Local and Integrated Forms of the Source Function: Definitions and Use 199
2.3 Which Other Reference Points, Beyond the Bond Critical Points? 203
3 The Source Function: Only a Mathematical Identity or Also an Interesting Chemistry Descriptor? 204
3.1 Chemical Transferability and the Source Function 205
3.2 Source Function Description of Simple and Well-Established Chemical Bonds 210
3.3 SF Description of Hydrogen-Bonded Systems 218
3.4 SF Description of Metal–Metal and Metal–Ligand Bonds in Organometallics 231
4 Is the Source Function a "Robust" Descriptor? 263
5 Double Integrating the Local Source: An Unambiguous Position-Space "Population Analysis" 268
5.1 Preliminary Results for the Second-Row Hydride Series and Comparisons with Mulliken's Population Analysis Scheme 269
6 Orbital and Core–Valence Decompositions of the Atomic SF Contributions 273
6.1 Alternative Decompositions of the Atomic SF Contributions and of Localization and Delocalization Indices 278
References 279

Abbreviations

±CAHB	Positively(negatively) charge-assisted hydrogen bond
Bcp	Bond critical point (in RFW Bader's theory)
BP	Bond path (in RFW Bader's theory)
Bza	Benzoylacetone
CC	Charge concentration
Cp	Critical point (in RFW Bader's theory)
DAFH	Domain-averaged Fermi hole
DFT	Density functional theory
ELF	Electron localization function
HB	Hydrogen bond
HF	Hartree–Fock
HS	Hirshfeld surfaces (M Spackman's definition)
IAM	Independent atom model
IAS	Interatomic surface (in RFW Bader's theory)
ICP	Interchanged population
IHB	Isolated hydrogen bond
IP	Ignored population
IQA	Interacting quantum atoms
LBHB	Low-barrier hydrogen bond
LS	Local source function

MM	Multipole model
MMED	Multipole model experimental density
MMPD	Multipole-modeled primary density
Mp	Midpoint (along an internuclear axis)
MPA	Mulliken's population analysis
NBO	Natural bond order
Nma	Nitromalonamide
PAHB	Polarization-assisted hydrogen bond
PD	Primary density (usually from ab initio computations)
QTAIM	Quantum theory of atoms in molecules (RFW Bader's theory)
RAHB	Resonance-assisted hydrogen bond
Rcp	Ring critical point (in RFW Bader's theory)
Rp	Reference point
SF	Source function
SSHB	Short-strong hydrogen bond
TMM	Trimethylenemethane complex
VSCC	Valence shell charge concentration (RFW Bader's theory)

1 Introduction

This chapter reviews a new chemical descriptor, first proposed by Bader and Gatti in 1998 [1], and whose peculiar properties are by nature profoundly germane to the main focus of this book relating *electron density and chemical bonding*. The source function enables one to view chemical bonding and other chemical paradigms from a totally new perspective and using only information from the electron density observable and its derivatives. It is completely independent from the tools used to obtain the electron density, which may be derived either through experimental techniques or with one of the many available quantum mechanical models at different levels of complexity. While the SF has been previously introduced in very concise sections of two general reviews [2, 3] and of a book chapter [4], this represents the first dedicated, comprehensive, and critical overview on the subject.

2 The Source Function: Basic Aspects

This section briefly summarizes the main aspects of the source function (SF), showing how it is mathematically derived and how it can be physically interpreted. Some necessary terminology is introduced, related to the different forms (local, integrated, or double integrated) this function may be used. Potential applications using one of these forms are also outlined.

2.1 Mathematical Derivation and Physical Interpretations

More than 10 years ago, Bader and Gatti showed [1] that the electron density at any point **r** within a system may be regarded as consisting of contributions from a local source, LS(**r**, **r**′), which operates at all other points of the space:

$$\rho(\mathbf{r}) = \int \mathrm{LS}(\mathbf{r}, \mathbf{r}') \cdot \mathrm{d}\mathbf{r}'. \tag{1}$$

The local source, given by $\mathrm{LS}(\mathbf{r}, \mathbf{r}') = -(4\pi \cdot |\mathbf{r} - \mathbf{r}'|)^{-1} \cdot \nabla^2 \rho(\mathbf{r}')$, where $(4\pi \cdot |\mathbf{r} - \mathbf{r}'|)^{-1}$ is a Green's function or an influence function [5], representing the effectiveness of how the cause, the Laplacian of the density at $\mathbf{r}'$, $\nabla^2 \rho(\mathbf{r}')$, gives rise to the effect, the electron density at **r**, $\rho(\mathbf{r})$. If the local source is integrated over an atomic basin or a group of atomic basins defined as in the quantum theory of atoms in molecules (QTAIM) [6], that is over those regions of space Ω bounded by surfaces S which are never crossed by $\nabla\rho(\mathbf{r})$ vectors $[\nabla\rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0 \ \forall \mathbf{r} \in S_\Omega]$, then the electron density at **r** may be equated to a sum of contributions $\mathrm{S}(\mathbf{r};\Omega)$, each of which is termed as the *source function* from the atom or group of atoms Ω to $\rho(\mathbf{r})$. When compared to $\rho(\mathbf{r})$, it represents a measure of how this atom or group of atoms Ω contributes to determine the density at **r**, relative to the contributions from other atoms or group of atoms in the system. One may visualize the electron density at a point within a given basin Ω as determined by an internal SF self-contribution and by a sum of SF contributions from the remaining atoms or groups of atoms within a molecule:

$$\rho(\mathbf{r}) = \mathrm{S}(\mathbf{r}, \Omega) + \sum\nolimits_{\Omega' \neq \Omega} \mathrm{S}(\mathbf{r}, \Omega'). \tag{2}$$

Decomposition afforded by (2) enables one to view the properties of the density from a new perspective and anticipates the SF as a tool able to provide chemical insight. Let us consider, for instance, the well-known and highly debated concept of bond path (BP) [7, 8] in Bader's theory. Topologically, a BP is associated with the only two atoms it connects [6], but its shape and the values of the electron density at any point along the path, including the so-called bond critical point (bcp), clearly depend, though to different extents, on the whole set of physical interactions present in a system and accounted for by its Hamiltonian operator. Using the SF, such an apparent inconsistency automatically fades away since, according to (2), the electron density at the bcp, taken as the most representative density point for the two bonded atoms [6], is determined not only from the contributions of these two atoms but also, in principle, from those of all the remaining atoms in the system, so bringing to the fore interesting "nonlocal" roles into the bonding. One immediately envisages that the more covalently will be two atoms bonded to each other, the higher will be their relative contributions to the density value at their intervening bcp [9].

Clearly, the integration on the whole space in (1) may be apportioned among the pieces of any conceivable mutually exclusive partitioning of space or in terms of one of the many proposed fuzzy boundary partitioning schemes [10]; however, since the QTAIM basins have the unique property of being proper open quantum systems and have also been amply demonstrated to be the atoms or group of atoms of "chemistry" [6], only such a specific partitioning may ensure an unbiased and quantum-mechanically rigorous association of $S(\mathbf{r};\Omega)$ with the atoms or group of atoms of a system.

A closer inspection to (1) and to the definition of the local source LS reveals [1] that $\rho(\mathbf{r})$ is given by an expression that resembles that for the electrostatic potential at $\mathbf{r}$ due to the electron distribution at all other points of the space, with $(4\pi)^{-1} \cdot \nabla^2\rho(\mathbf{r}')$ replacing $\rho(\mathbf{r}')$ in the numerator of the integrand. Indeed, both the electrostatic potential V_{elec} and $\rho(\mathbf{r})$ are a solution [11]

$$\varphi(\mathbf{r}) = \int (4\pi \cdot |\mathbf{r} - \mathbf{r}'|)^{-1} \cdot q(\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{3}$$

of the Poisson's equation $\nabla^2\varphi(\mathbf{r}) = -q(\mathbf{r})$, with φ being, respectively, V_{elec} or ρ, q being, respectively, ρ or $\nabla^2\rho$, and exploiting the definition of V_{elec} in terms of Poisson's equation, $\nabla^2 V_{\text{elec}}(\mathbf{r}) = +4\pi \cdot \rho(\mathbf{r})$.[1] The electron density $\rho(\mathbf{r})$ may thus be seen as the potential generated by its Laplacian distribution [1], in agreement with the physical interpretation given earlier as of $\nabla^2\rho(\mathbf{r}')$ causing or determining $\rho(\mathbf{r})$.

It has been recently claimed that [12] one should focus more on the physical interpretation of the SF, namely that the Laplacian distribution *determines* the electron density at any point in space, rather than on the "formal" mathematical interpretation of (2) that a basin *contributes* to the density. Although we believe not be responsible of any serious misconception in this regard, it seems yet worth spending few words to further clarify the point. Since the Laplacian distribution *determines* ρ at any point in space, integration of such a distribution within an atomic basin, weighted by the influence function, just singles out the distinct *contribution* from the basin to the *determination* of ρ.[2] It is in this sense that, within the SF approach, an atom gives its contribution to $\rho(\mathbf{r})$, and not clearly in terms of the direct contributions from its basis functions to the total density, even admitting one could always define them (which is not the case, for instance, for electron densities given numerically on a grid). An interpretation of the SF (and of its integrated form, see *infra*) in terms of a standard population analysis would also be totally at odd with the exhaustive and mutually exclusive partitioning of the space adopted within the SF approach.

[1]The right-hand member has a positive sign since the electron density $\rho(\mathbf{r})$ is taken as a positive quantity, despite the electron is negatively charged.

[2]Incidentally, one should note that a uniform distribution has no sources, since $\nabla^2\rho$ would vanish everywhere in this case and the only source for the density at point $\mathbf{r}$ is the point itself.

Use of the local expression of the virial theorem [6]

$$\frac{1}{4}\cdot\nabla^2\rho(\mathbf{r}) = 2\mathrm{G}(\mathbf{r}) + \mathrm{V}(\mathbf{r}), \tag{4}$$

enables one to express the local source in terms of the positively defined kinetic energy density, G(**r**), and of the electronic potential energy or virial density, V(**r**), and to so introduce [1] a further interesting interpretation of the local source

$$\mathrm{LS}(\mathbf{r},\mathbf{r}') = -\frac{1}{\pi}\cdot\frac{2\mathrm{G}(\mathbf{r}) + \mathrm{V}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} = \mathrm{LG}(\mathbf{r},\mathbf{r}') + \mathrm{LV}(\mathbf{r},\mathbf{r}'). \tag{5}$$

Equation (5) discloses that LS is related to the failure to locally satisfy the virial relationship between twice the integrated kinetic energy and the virial field densities. It also shows that LS may be seen [13] as given by the sum of a kinetic energy, LG, and of an electron potential energy, LV, local source contribution.[3] Molecular regions, where the electron density is concentrated ($\nabla^2\rho(\mathbf{r}') < 0$) and where the potential energy dominates over the kinetic energy, act as a positive *source* for the electron density at a point **r**, whereas regions of charge depletion, ($\nabla^2\rho(\mathbf{r}') > 0$), and of dominant kinetic energy act as a negative source, a *sink*, for the same point. The effectiveness of the electron density at $\mathbf{r}'$ to be source or sink for the electron density at another point **r** is then related to the magnitude of its charge concentration or depletion at $\mathbf{r}'$, weighted by the inverse of the distance of these two points.

A given atomic source function value, S(**r**;Ω), will always be the result of the sum of local positive and negative contributions and can thus be either globally positive or negative. As $\rho(\mathbf{r})$ is positive everywhere, S(**r**;Ω) will also be positive at any **r** for an isolated atom, since its own basin is the only one determining the density. For an atom in a polyatomic system, the local sources are usually found to beat the local sinks in determining the electron density at its intervening bcps, but the opposite may also hold true in specific circumstances. A typical example is the source from the hydrogen atom involved in standard hydrogen bonds to the electron density at the hydrogen bcp (see Sect. 3.3).

This section ends by introducing a second powerful formula for the density at a point in terms of external sources and by showing how both (1) and this formula may easily be derived from a single mathematical theorem. The alternative expression for $\rho(\mathbf{r})$ is given by [1]

$$\rho(\mathbf{r}) = -(1/4\pi)\cdot\left\{\int_\Omega \frac{\nabla^2\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + \oint_{S_\Omega} dS(\mathbf{r}_S)\bullet\nabla|\mathbf{r}-\mathbf{r}_S|^{-1}\rho(\mathbf{r}_S)\right\}, \tag{6}$$

[3] LS could also be expressed in terms of local contributions related to the total electronic energy density H(**r**) [6] and to the kinetic energy density G(**r**) since ¼ $\nabla^2\rho = 2\mathrm{G}(\mathbf{r}) + \mathrm{V}(\mathbf{r}) = \mathrm{G}(\mathbf{r}) + \mathrm{H}(\mathbf{r})$.

where the density at a point **r** within an atom Ω is seen as determined by the sum of two contributions: the basin average of the potential at **r** exerted by the Laplacian of the density (the basin average of the local source LS), plus the flux through the surface of Ω of the electric field density at **r** due to the electron density on the surface boundary, $\rho(\mathbf{r}_s)$.[4] For an isolated molecule or a molecular complex, Ω may be clearly taken as the whole space and (6) becomes identical to (1), as the electron density vanishes at infinity. Formula given in (6) should be applied to those systems which have finite boundaries, like a (unit) cell in a crystal or any of its composing atoms or group of atoms. It should also be used for large molecular systems, or for convenient and well-defined part of them. One could so replace the integration of the LS over all (potentially infinite) atoms of the system, with just one basin's and one surface's integration. For instance, the effect of the environment on a cluster of molecules within a liquid or crystal might be conveniently investigated, through the SF approach, using the surface boundary of the cluster in (6). The surface integral can then be envisaged as a sum of contributions from each of the interatomic surfaces $S(\Omega,\Omega')$ composing the total surface $S(\Omega)$ bounding Ω.

As shown in [1], (6) may be obtained in several ways: through the use of the equation of motion for the generator $|\mathbf{r}-\mathbf{r}'|^{-1}$for a proper open system Ω (a QTAIM basin), or by solving Poisson's equation for a potential given by the density $\rho(\mathbf{r})$, or also in a purely mathematical manner, using the Green's theorem

$$\int_{\mathrm{A}}(u\nabla^2 v - v\nabla^2 u)\cdot \mathrm{d}\mathbf{r}' = \oint_{S_\mathrm{A}} \mathrm{d}S\cdot(u\nabla v - v\nabla u), \tag{7}$$

with A being an arbitrary basin and S_A its enclosing surface. By making use of the well-known [5] identity, $\nabla^2(|\mathbf{r}-\mathbf{r}'|)^{-1} = -(4\pi\cdot\delta(\mathbf{r}-\mathbf{r}')$ and by setting $u = |\mathbf{r}-\mathbf{r}'|^{-1}$ and $v = \rho(\mathbf{r}')$, one easily gets (6) *provided* the basin A fulfills the QTAIM zero-flux recipe $[\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 0 \quad \forall\mathbf{r}\in S_\mathrm{A}]$, to get rid of the surface term $u\nabla v$.

2.2 *Local and Integrated Forms of the Source Function: Definitions and Use*

In Sect. 2.1, the local form of the source function, (8)

$$\mathrm{LS}(\mathbf{r},\mathbf{r}') = -(4\pi\cdot|\mathbf{r}-\mathbf{r}'|)^{-1}\cdot\nabla^2\rho(\mathbf{r}'), \tag{8}$$

[4]In fact, $\nabla(|\mathbf{r}-\mathbf{r}_s|)^{-1}\,\rho(\mathbf{r}_s) = -(\mathbf{r}-\mathbf{r}_s)/|\mathbf{r}-\mathbf{r}_s|^3\,\rho(\mathbf{r}_s) = \varepsilon\,(\mathbf{r},\mathbf{r}_s)$, with $\varepsilon\,(\mathbf{r},\mathbf{r}_s)$ being the electric field density at **r** due to the electron density at $\mathbf{r}_s$.

and its integration over a basin Ω

$$S(\mathbf{r},\Omega) = \int_{\Omega} LS(\mathbf{r},\mathbf{r}') \cdot d\mathbf{r}', \tag{9}$$

were introduced and termed, respectively, *Local Source* (LS) function and (integrated) *Source Function* (SF) for the electron density at **r**. They express the contribution in determining $\rho(\mathbf{r})$ of a local source at $\mathbf{r}'$ or of the sum of this source at all points within Ω, respectively.

A standard use of the SF approach [1, 9, 14] involves the evaluation of the SF contribution of an atomic basin or of a convenient union of atomic basins (e.g., a chemical group) in determining the density at **r**, chosen as a reference point (rp). Typically, the bcp is used as the least biased choice for the point representative of a bonding interaction. Analogously to the calculation of the atomic population N(Ω), integration of the LS is usually performed using nucleus-centered spherical coordinates. Therefore, the SF contribution from a given group of atoms is conveniently performed [1, 9, 14] by summing up the individual S(**r**,Ω) contributions of the various atoms composing the group, rather than by a direct integration over the basin of the whole group.

The SF values, S(**r**,Ω), are often reported [9, 14] in terms of their percentage contribution to the electron density at **r**,

$$S\%(\mathbf{r},\Omega) = \left[\frac{S(\mathbf{r},\Omega)}{\rho(\mathbf{r})}\right] \cdot 100. \tag{10}$$

It is worth noting that the S(**r**,Ω) and S% (**r**,Ω) quantities have a quite distinct meaning. For instance, in the case the rp corresponds to a bcp, S(**r**,Ω) is strictly related both to the nature and strength of the associated interaction, as conveyed by the electron density at bcp, ρ_b, and by the relative "contribution" from Ω to ρ_b. On the other hand, S%(**r**,Ω) just expresses the percentage share from Ω to ρ_b, and it is therefore in principle independent on how strong or weak the interaction is. S(**r**,Ω) may be very small or large, despite S%(**r**,Ω) being, respectively, very close to and even larger than 100%[5] or definitely much less than so. As we will show S%(**r**,Ω) reflects the localized vs. delocalized character of a given chemical interaction, with the two bonded atoms having large S%(**r**,Ω) values when the interaction is localized and quite small when highly delocalized. Both quantities, S(**r**,Ω) and S%(**r**,Ω), will be used and analyzed in the examples discussed in the next sections.

With respect to its integral forms, the local SF brings further detail. The LS is currently analyzed [13] by evaluating its profile along a line and using a convenient point as rp. When the line is a bond path, the bcp is usually taken as the rp.

[5]Negative S(**r**,Ω) values are not uncommon and S%(**r**,Ω) may thus be negative, which implies that the percentage contributions from other atoms may occasionally become greater than 100.

In this way, atomic regions yielding positive and negative contributions to ρ_b become clearly visible along the bond [13, 14] and may be associated with the corresponding regions of charge concentration and charge depletion as obtained from the radial profile of $\nabla^2\rho$ in the isolated atom [6]. Comparison of LS profiles for the same system and for the same rp, but using electron densities coming from different models, allows one to get a deeper understanding of the changes brought in by the various adopted models into the description of bonding at the bcp [13–15]. Comparison of LS profiles may also disclose interesting differences in how the electron density is determined at a bcp or at this same location when two atoms are linked or not linked by a bond path. Examples of such an use of the LS profiles are detailed in Sects. 3.4.2 and 4, where the application of the SF tool to the study of metal–metal bond and of the multipole bias in charge density investigations from X-ray structure refinements is, respectively, discussed.

Analogously to the LS profile along a line, one may also investigate $S(\mathbf{r},\Omega)$ or $S\%(\mathbf{r},\Omega)$ along this same line [13, 14] (see Sects. 3.1 and 3.4.2). The resulting profile clearly bears a quite different interpretation, since in this case, at variance with the LS profile, it is the rp that is changing along the line and not the LS contributing point. The $S(\mathbf{r}, \Omega)$ or $S\%(\mathbf{r}, \Omega)$ profiles thus give information on how the total or percentage "contribution" from Ω to $\rho(\mathbf{r})$ varies as a function of the position of $\mathbf{r}$ along the line. If the line is a bond path linking Ω and Ω', the "contribution" from Ω' will generally increase on going from Ω to Ω'. Participation from other atoms will also vary, with a generally larger importance in regions around the bcp and far from either of the two linked nuclei. The relative small or large weight of contributions from atoms other than Ω or Ω', in the region around the bcp, may be respectively related to the high or low covalent character of the interaction Ω–Ω'.

Although only partly explored thus far [3, 16], the SF tool may also be used to define the extent to which a basin contributes to determine both its own electron population $N(\Omega)$ and that of the remaining basins in the system. By integrating (2) over Ω

$$N(\Omega) = \int_{\Omega} \rho(\mathbf{r})d\mathbf{r} = \int_{\Omega} S(\mathbf{r},\Omega)d\mathbf{r} + \sum_{\Omega' \neq \Omega} \int_{\Omega} S(\mathbf{r},\Omega')d\mathbf{r} = N_i(\Omega) + N_o(\Omega), \quad (11)$$

or, analogously, (6)

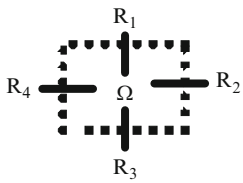
$$\begin{aligned} N(\Omega) &= \int_{\Omega} \rho(r)dr \\ &= -\frac{1}{4}\pi \cdot \int_{\Omega} \left\{ \int_{\Omega} \frac{\nabla^2\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + \oint_{S_\Omega} dS(\mathbf{r}_S) \cdot \nabla|\mathbf{r}-\mathbf{r}_S|^{-1}\rho(\mathbf{r}_S) \right\} d\mathbf{r} \\ &= N_i(\Omega) + N_o(\Omega), \end{aligned} \quad (12)$$

$N(\Omega)$ turns out to be decomposable in an inner, $N_i(\Omega)$, and an outer, $N_o(\Omega)$, contributions, which in the case of (11) may be then both conveniently expressed in terms of the elements of an electron "population" matrix M.

$$N(\Omega) = N_i(\Omega) + N_o(\Omega) = M(\Omega, \Omega) + \sum_{\Omega' \neq \Omega} M(\Omega, \Omega'). \tag{13}$$

$M(\Omega,\Omega)$ represents the contribution of Ω to determining the electron population of its own basin, while $M(\Omega,\Omega')$ is the contribution to this same population from basin Ω'. $M(\Omega',\Omega)$, instead, represents the contribution of Ω in determining the population of Ω' and, usually, $M(\Omega,\Omega') \neq M(\Omega',\Omega)$. At variance with standard population analyses, the matrix M is thus generally not symmetric, although it will be clearly so in specific cases (e.g., the homonuclear diatomics). In the case of (12), the outer contribution would consist of a sum over the interatomic surfaces of Ω of the global contributions from each chemical residue R linked to Ω (Scheme 1).

The matrix M (13) defines a full "population analysis," based only on the observable $\nabla^2\rho$ and without resorting, as it is generally the case, to any discrete representation of the one electron density matrix on atomic bases and to population analyses derived thereof (e.g., Mulliken Population analysis, MPA) [17]. As anticipated earlier, the quotes enclosing "population analysis" signify the profound difference between an atomic population given in terms of a sum of contributions to the density from the atomic basis set functions (e.g., MPA) and the present one given in terms of contributions from well-defined portion of spaces *determining* that population. For the sake of simplicity, we will, however, omit the quotes from now on when referring to either type of populations. The matrix M, analogously to the local source and to the source function, can be obtained from experimental electron densities. It thus represents a unique chance to compare theoretical and experimental electron populations on an unbiased basis, regardless of the different descriptions – atomic basis sets, pseudoatom densities [10], etc. – adopted to derive them. It is worth noting the formal analogy between the proposed population analysis and that one would obtain by defining the boundaries of atoms according to the Hirshfeld surfaces (HS) introduced by Spackman and Byrom [18], and by then using the so-called Hirshfeld's stockholder partitioning [19] to apportion, among the various atoms, the electron density within the HS bounded basins. This procedure would automatically lead to a nonsymmetric M matrix formally analogous to that obtained by the double integration of the local source and would equally be applicable, on the same grounds, to both theoretical and experimental densities. An interesting aspect



Scheme 1 Atomic group Ω and its linked chemical residues R

is that this population analysis would be definitely much faster and easy to calculate, but it also has significant drawbacks respect to that based on the source function. First, the atomic boundaries and the apportioning among atoms of the density within them would be biased by the independent atom model (IAM) [10] approach, which is arbitrary and also not unique. Second, the so-defined atoms would not be quantum objects. Third the HS space partitioning is not necessarily exhaustive, which would imply the existence of basins – the voids of usual HS studies [18] – not associated with any nucleus. In any case, the analogy between these two populations would be only formal since the one making use of HS surfaces and the stockolder recipe is defined, as usual, in terms of direct contributions to the density, while that based on the double integration of the local source, in terms of contributions *determining* the density. Preliminary results and numerical difficulties and problems using (11)–(13) are reported in Sect. 5.1.

2.3 *Which Other Reference Points, Beyond the Bond Critical Points?*

Before concluding this general introduction to the LS and the SF, it is worth spending few words on the choice of the more relevant reference points when using these functions. Needless to say this problem does not come up when the LS is used in its double integrated form.

As discussed above, both the LS and the SF often use bcps as reference points, but clearly any reasonable choice may be exploited and proved to be useful. The case of the SF profile along a bond path has already been introduced earlier. Among the many other possible and convenient choices, we mention (a) the positions of maximum local charge concentrations or depletions [6], so as to explore in an atomic-wise detail how their Laplacian values are affected by chemical substitution, and (b) the points along a line perpendicular to the bond path and directed above and below the molecular σ-plane in a π-conjugated system. This latter option is expected to provide information on how π-conjugation is mirrored by the SF. In fact, even though σ- and π-distributions are well known to be self-consistently interrelated [6, 20], π orbitals do not yield *direct* contributions to the electron density in the molecular σ-plane for those systems where symmetry allows for a complete σ and π separation. Hence, one anticipates an increase in the S% contribution from atoms other than the two linked by a bond path when the relative contribution from π-orbitals to the total density rises on moving out from the molecular plane. An example of such behavior is shown in Sect. 3.2.2.

As we discussed in [21], in reply to a question by Professor Bultinck, use of the ring critical points (rcp) [6] in aromatic systems as rp's could reveal different S% (rcp,Ω) contributions from the various atoms in the ring and so provide a measure of departure, if any, from the perfect atomic symmetry necessarily present in benzene. Such a departure, although not yet extensively analyzed, might be used to define an alternative (local) index of benzenoid character [22].

The points associated with the attractors of nonnuclear basins [23, 24], that is the maxima of $\rho(\mathbf{r})$ at positions other than nuclei, have also been used [9] as rp's so as to explore whether the SF is able to distinguish between these peculiar maxima, with anticipated highly delocalized character for their density sources, and those associated with the nuclei, presumably characterized by very localized sources. Gatti et al. [9] surveys the results obtained by applying such an analysis to a number of alkali metal clusters exhibiting nonnuclear basins in their electron distribution.

Other rp's choices are clearly possible besides those listed, but a discriminating point needs to be mentioned [14]. The relative accuracy by which $\rho(\mathbf{r})$ is obtained as a sum of atomic sources, (2), is given [14] by the quantity ER%($\mathbf{r}$)

$$\mathrm{ER}\%(r) = \left| \frac{\rho(\mathbf{r}) - \sum_{\Omega} \mathrm{S}(\mathbf{r}, \Omega)}{\rho(\mathbf{r})} \right| \cdot 100, \tag{14}$$

which turns out to be usually less than 1% when $\rho(\mathbf{r})$ is greater than 10^{-2} au, but which may increase up to about 5% when $\rho(\mathbf{r})$ is one order of magnitude lower. For regions of very low density, $\rho(\mathbf{r}) < 10^{-4}$ au, reconstruction of $\rho(\mathbf{r})$ through (2) often becomes problematic. The modulus of the sum of atomic sources $\left|\sum_{\Omega} \mathrm{S}(\mathbf{r}, \Omega)\right|$ hardly becomes lower than 10^{-5} au, or, in some case, even than 10^{-4} au, which easily explains why *ER%*($\mathbf{r}$) is rapidly increasing up to and above 100% for rp's with very low density values. This accuracy problem clearly sets a minimum density value constraint ($\approx 10^{-3}$ au) on the choice of possible rps. It has also been shown [14] that the ability of reconstructing the density through (2) is strictly related to the departure of the integrated atomic Laplacian magnitudes, $\left|\int_{\Omega} \nabla^2 \rho(\mathbf{r}) \cdot \mathrm{d}\mathbf{r}\right|$, from their required value of zero. For instance, in the isolated H atom, with an integrated Laplacian magnitude less than 10^{-11} au, $\rho(\mathbf{r})$ values as small as 10^{-10} au are found to be reconstructed with ER% less than 10%, while for the H_2 molecule, with an integrated Laplacian of about 10^{-5} au, ER% values become already noticeable for $\rho(\mathbf{r})$ values as large as 10^{-4} au. More details on the problem of reconstructing the electron density from atomic sources and on the numerical techniques which have been devised to increase the overall reconstruction accuracy are discussed in [14].

3 The Source Function: Only a Mathematical Identity or Also an Interesting Chemistry Descriptor?

This section enters into the realm of the source function applications to chemistry. The main concern we had [1, 9] since the first proposal of such a function in 1998 was to explore whether (2) was indeed able to provide chemical insight or whether it represented just a mathematical identity for a *trivial* "tautological reconstruction

of the electron density" [12]. The physical interpretations of the source function discussed in the previous section foresee this function as capable to mirror in some way the effects that typical *chemistry changes*, such as chemical bonding, chemical substitution, and chemical environment, or *paradigms*, such as chemical transferability and the notion of chemical groups, bring into the electron density. We challenge, in the following, whether such an anticipated ability is warranted.

We start by considering how the SF may provide an innovative and peculiar view of how transferability realizes in chemistry; we next move to the SF analysis of very simple and well-established prototypical chemical bond patterns so as to get some feeling on how the SF describe and possibly distinguish their different nature. Later, hydrogen-bonded systems and metal–metal or metal–ligand interactions in organometallics are inspected to challenge the information the SF is able to provide in less conventional bonding cases.

3.1 Chemical Transferability and the Source Function

Usually, chemical transferability of a given piece of matter – e.g., a group of atoms or a molecule – is examined in terms of the constancy, to a given extent, of a number of its properties despite the different chemical environment in which it is placed. *Perfect transferability* is achieved when the electron density of such a piece of matter is fully transferable [6], while (partial) transferability of only some of its properties may realize through what has been termed *compensatory transferability* [25, 26]. This might be for instance the case of a constant electron population for an atomic group, which realizes either because of a compensation of charge transfers within the atoms of the group, or even through self-charge polarization mechanisms within one or more atoms of the group. The group population remains the same, but the electron density of the group is not fully transferable in such a case.

The source function is of use in determining individual group contributions to the density in the study of transferability and may also serve to reveal the consequences of the transferability of the properties of a functional group, since, as said earlier, the extent of transfer of these properties from one molecule to another is a consequence of a corresponding transferability of the group's electron density. For instance, the SF enables one to determine the extent to which changes in atoms neighboring the group in question selectively contribute to the change in its density.

Using the SF, one may also see transferability from a *new and deeper perspective*. The electron density decomposition afforded by (2) reveals that the perfect transferability of a group property, expressible in terms of its density, implies a constancy not only in the electron density of the group but also in the sum of contributions to this density from the remaining atoms in the system. Concisely, if on passing from one system to another, a group's electron density remains constant, so need to be both the "internal" and "external" contributions to this density, at any of its points.

In what follows, we briefly review a case where an almost perfect transferability realizes and other two where the SF approach reveals interesting compensatory transferability mechanisms.

Richard Bader has repeatedly shown how transferability of form and properties is particularly evident for atomic groupings corresponding to the building block of biological macromolecules [25, 27] or of some series of molecules, like the hydrocarbons [6]. In particular, the terminal methyl group in *n*-alkanes, past ethane, is characterized by very transferable atomic properties, regardless of the length of the chain, and including its energy, electron population, volume, and spectroscopic behavior [6]. Transferability does not only show up in the integrated properties, but it is so good that, for instance, a constant value of 0.2827 au for ρ_b at the unique C–H bond of the terminal methyl group is also detected (RHF/6-311G** level). Analysis of SF contributions to this density value in the ethane, propane, butane, and pentane series leads [1], Table 1, to two important results: (a) the "internal" contributions from the atoms in the methyl group are actually constant at 0.270 au throughout the series, with the two equivalent methyl group hydrogen atoms contributing 0.0210 au in all four molecules, and (b) the "external" contribution remains also almost constant, regardless of the length of the chain. The two SF conditions of a *perfect transferability* are thus both fulfilled for the methyl group, at least at its C–H bcp. The constancy of the external contribution implies that the ethyl group in propane has to contribute the same as the propyl group in butane or the butyl group in pentane. It is this constraint that leads the H and C atoms of the methyl group to exhibit characteristic properties in hydrocarbons. The trend

Table 1 Source function (SF) and atomic group transferability. Source contributions to the H–CH_2 bcp density in $CH_3(CH_2)_nCH_3$ (n = 0–3) and to the Li–*X* bcp density in the Li–*X* (*X* = F, O, N, Cl, H) Series[a]

n-hydrocarbons, $CH_3(CH_2)_nCH_3$							
N	ρ_b(H–CH_2)	S($\mathbf{r}_b$; ext)[b]	S($\mathbf{r}_b$, Ω)[b]				
0	0.2830	0.0126	H–CH_2·····CH_2·····		H		
			0.2704	0.0100	0.0026		
1	0.2827	0.0126	H–CH_2·····CH_2·····		CH_3		
			0.2701	0.0091	0.0035		
2	0.2827	0.0127	H–CH_2·····CH_2·····		CH_2·····	CH_3	
			0.2701	0.0091	0.0020	0.0016	
3	0.2827	0.0127	H–CH_2·····CH_2·····		CH_2·····	CH_2·····	CH_3
			0.2702	0.0090	0.0019	0.0008	0.0009
Li–X series							
Li–*X*	[R_e (R_{Li})][c]	ρ_b	S($\mathbf{r}_b$, Li)	S%($\mathbf{r}_b$, Li)	N(Li)[d]	−E(Li)[d]	
Li–F	2.935 (1.128)	0.078	0.033	42.9	2.059	7.3419	
Li–O ($^2\Pi$)	3.158 (1.175)	0.067	0.028	41.6	2.066	7.3467	
Li–N ($^3\Sigma$)	3.477 (1.242)	0.054	0.022	39.9	2.075	7.3565	
Li–Cl	3.846 (1.294)	0.045	0.018	39.8	2.065	7.3274	
Li–H	3.039 (1.347)	0.040	0.016	40.0	2.086	7.3655	

[a]Data from [1, 3, 9], all quantities in au

[b]S($\mathbf{r}_b$; ext) is the sum of "external" source contributions to ρ_b (H–CH_2), i.e., those other than from the methyl group. It is thus given by the sum of the atomic group source contributions listed, for each molecule, in the last column minus S($\mathbf{r}_b$, Ω = H–CH_2)

[c]R_e and (R_{Li}) are the equilibrium distance and the distance from the Li nucleus to the bcp

[d]N(Li) and E(Li) are the electron population and the atomic energy of Li

reported in Table 1 is much more than a simple falloff in the source contributions with chain length increase. The value of the source function for each succeeding increment is predetermined as it must equal the contribution of the hydrogen atom it replaces.

In our second example [9], we examine the Li–X (X = F, O, N, Cl, H) diatomics where the Li atom is known to exhibit almost constant and transferable integral properties, including its net charge and atomic energy, through the series [6, 28]. Before introducing the SF view of transferability along this series, it is worth remembering that the charge distributions of Li–H, Li–O, and Li–F were used back in 1972 by Bader as the examples to "illustrate the all important observation that the transferability of atom's density is accompanied by a paralleling transferability in its kinetic energy density" [28]. Indeed, the paralleling behavior of $\rho(\mathbf{r})$ and G($\mathbf{r}$) was the "crucial observation" that eventually led to the QTAIM [6], next the demonstration [29] that the virial theorem not only holds for a system in its entirety, but also applies to each of its regions of space bounded by a zero-flux surface in $\nabla\rho$. Constancy of Li integral properties also serves as a simple example of the most important observation at the basis of the concept of a functional group: "that atoms or linked groupings can exhibit characteristics forms and properties in spite of gross changes in their immediate neighbors" [25].

In the limit of perfect transferability of the Li basin electron distribution, the source contribution from Li to the density at the bcp, S($\mathbf{r_b}$,Li), should remain constant, regardless of the nature of X and of the corresponding value of ρ_b [9]. Within the limit above, changes in the ρ_b value should only be determined by corresponding changes in the source contribution from the X atom, S($\mathbf{r_b}$,X). Table 1 shows, instead, that the S($\mathbf{r_b}$,Li) value provides a more sensitive measure of departure from perfect transferability of the Li basin than its integral properties. S($\mathbf{r_b}$,Li) is found to decrease with decreasing electronegativity through the series and to become in LiH about half the value in LiF, the first member of the series. The decrease in S($\mathbf{r_b}$,Li) parallels the corresponding decrease in the ρ_b values, whereas both the atomic electron population N(Li) and the energy E(Li) remain almost unvaried through the series. Interestingly, an almost constant source percentage contribution from Li at the bcp density, S%($\mathbf{r_b}$,Li) $\approx$ 40%, is also observed (Table 1). The constancy in Li energy and net population through the series is translated in the SF language in a constancy of the percentage share of the electron density at the bcp, rather than in an unaffected SF contribution. The chemical cationic nature of Li in the series thus implies a constant share of the bcp density and not a constant ρ_b value, which also confirms how the information provided by S($\mathbf{r_b}$,Ω) and S%($\mathbf{r_b}$,Ω) may substantially differ in many cases, as anticipated earlier.

The source function is also able to reveal that a mechanism of compensatory transferability operates within the Li basin to ensure constant integral properties. These latter are preserved in the series (Table 1) through a noticeable basin expansion toward X with decreasing X's electronegativity. Figure 1 shows that the S($\mathbf{r}$,Li) profiles along the Li–X axis for the various members of the series coincide up to a distance from the Li nucleus almost equal to that of the closest bcp to Li (LiF, R_{Li} = 1.128 au) and then start to only slightly differ after this

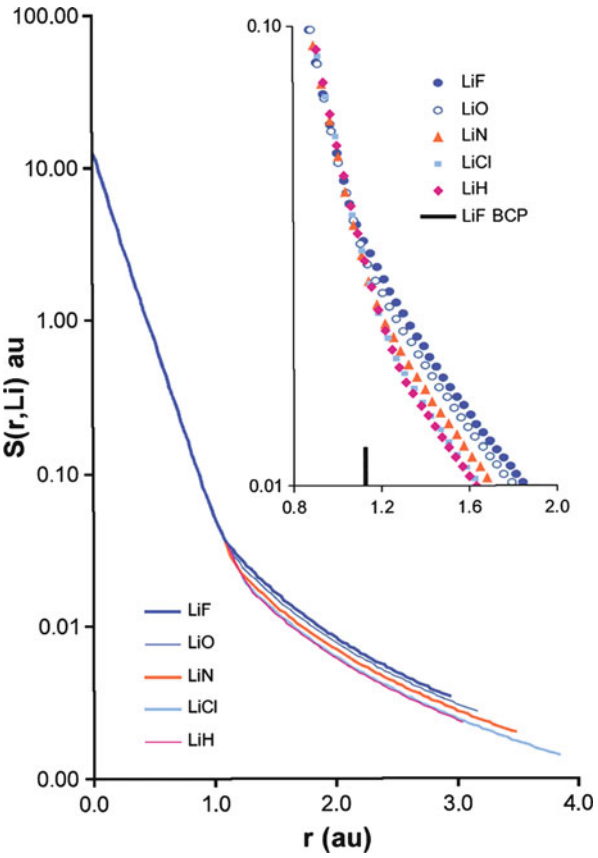


Fig. 1 Profile of source contributions from the Li atom, S(**r**, Li), to the electron density along the internuclear Li–*X* (*X* = F, O, N, Cl, H) axis. The Li nucleus is at $r = 0.0$. Source function contributions in the region of the bcps are enlarged in the inset where the vertical bar denotes the position of the LiF bcp, which is the closest to the Li nucleus. The observed changes in S(**r**, Li) with *X*, despite the almost constant Li electron population, are due to the bcp shift toward *X* through the series (reproduced from Fig. 3 with permission from [9], Copyright 2003, Wiley-VCH Verlag GmbH & Co, KGaA)

distance. It is the shift in the bcp position along the series and not the small difference in the S(**r**,Li) profiles which lead to the observed trend in the S($\mathbf{r}_b$,Li) values and which ensures a constancy in S%($\mathbf{r}_b$,Li) and in N(Li). The decrease in S(**r**,Li) with increasing distance from the Li nucleus is clearly evident from Fig. 1. It results from the increasing weight of the positive Laplacian regions when the rp moves toward *X* [9].

The third example on transferability reviewed in this section concerns the series of heteromolecules HCH_2–CH_2B with B = H, CH_3, NH_2, OH, and F [1]. Energies of this series of molecules are found to equal the arithmetic mean of the energies of the corresponding homomolecules HCH_2–CH_2H and BCH_2–CH_2B to within a few Kcal/mol or less, computationally and experimentally [1, 30]. These additivity rules suggest that the groups $HCH_2|$ and $BCH_2|$ (with vertical bars indicating the zero-flux surface of the C–C bond) undergo small changes in forming the heteromolecule. The largest density perturbations, when HCH_2–CH_2B is formed, should occur in the vicinity of the new C–C interatomic surface and in particular at its highest density value ρ_b. Table 2 reports ρ_b values at the C–C bcp and their departures, $\Delta\rho_b$,

Table 2 Source function (SF) and compensatory transferability. Source contributions to the C–C bcp in the H–CH_2–CH_2B (B = CH_3, NH_2, OH, F) series[a,b]

B	ρ_b	$(\rho_b)_{av}$ ($\Delta\rho_b$)	$\Delta S(H–CH_2)$[c]	$\Delta S(CH_2–B)$[c]	$\Delta S(C)$[d]	$\Delta S(H_2)$[d]
H	0.2514	–	–	–	–	–
CH_3	0.2532	0.2532 (0.0000)	0.0006	−0.0006	0.0007 (−0.0008)	−0.0002 (0.0001)
NH_2	0.2592	0.2581 (0.0011)	0.0044	−0.0033	0.0040 (−0.0032)	0.0002 (0.0000)
OH	0.2651	0.2632 (0.0019)	0.0078	−0.0059	0.0067 (−0.0047)	0.0004 (−0.0005)
F	0.2688	0.2652 (0.0036)	0.0108	−0.0072	0.0092 (−0.0060)	0.0007 (−0.0005)

[a]Data from [1], all quantities in au

[b]$(\rho_b)_{av}$ is the mean of the ρ_b value for C–C bond in HCH_2–CH_2H and in BCH_2–CH_2B, while $\Delta\rho_b$ represents the deviation from the mean value, $\Delta\rho_b = \rho_b - (\rho_b)_{av}$

[c]$\Delta S(H–CH_2)$ and $\Delta S(CH_2–B)$ are the source differences for these groups in H–CH_2–CH_2B and in the corresponding homomolecules, HCH_2–CH_2H and BCH_2–CH_2B

[d]$\Delta S(C)$ and $\Delta S(H_2)$ are the changes in the source contributions from the C and H_2 in the H–CH_2 group with respect to the values for the same group in HCH_2–CH_2H. In parenthesis, the corresponding changes in the CH_2B group with respect to the values for the same group in BCH_2–CH_2B

from the arithmetic mean of the homosubstituted values. The $\Delta\rho_b$ are very small relative to ρ_b and increase with the electronegativity of B. The SF analysis reveals that the small $\Delta\rho_b$ values are the result of much larger perturbations which almost compensate each other to yield very small deviations from the arithmetic mean additivity rule. $\Delta S(H–CH_2)$ and $\Delta S(CH_2–B)$ in Table 2 represent the deviations of source contributions to C–C ρ_b from the $HCH_2|$and $BCH_2|$groups in H–CH_2–CH_2B with respect to the values of these sources in the corresponding homomolecules. Their values are almost equal in magnitude but have a different sign and are from three to five times larger in magnitude than $\Delta\rho_b$. Table 2 also shows that both $\Delta S(H–CH_2)$ and $\Delta S(CH_2–B)$ are dominated by the changes experienced by the carbon atoms involved in the investigated C–C bond. The sources from the $HCH_2|$ and $BCH_2|$ groups, and the corresponding sources from the C atoms of the C–C bond, are found to, respectively, increase, relative to ethane, and decrease, relative to B–CH_2–CH_2B, by similar amounts so as to yield the final small $\Delta\rho_b$ values.

Although the discussed results seem quite promising, application of the SF analysis to the study of transferability is certainly still in its infancy. The SF capability to reveal and detail external, possibly long-distance contributions in determining the electron density and its changes at relevant points, upon chemical substitution or perturbation by the environment, might become of some interest in pharmacological and biological studies. For instance, as an interesting tool for drug design or in the analysis of how close to the actual drug/receptor interactions are the models taken from their simplified and perturbed interactions in crystal assemblies [31]. A preliminary work along these directions, aimed at exploring whether the SF descriptor is able to finely disentangle the effect of the environment on the charge distribution of inhibitor species, is included in the chapter by Bernd Engels et al. [32] of the present book.

3.2 *Source Function Description of Simple and Well-Established Chemical Bonds*

This section reviews the use of the SF for studying chemical bond features, starting from few very simple and well-known chemical bonds. We show how the SF describes what has been already largely investigated using either the molecular orbital approach or one of the various available topological tools. As anticipated earlier, exploring a further method like the SF is primarily justified by its possible application, on the same ground, to both experimental and theoretical densities, and second by the reasonable curiosity to inspect whether such a peculiar method is able to provide any chemical insight.

3.2.1 CH_n–CH_n (n = 1–3) Hydrocarbons, B_2H_6, BH_3, BH_3–PH_3

Figure 2 shows the SF contributions at the bcp for a number of compound with either increasing covalent bond order (the C–C bonds in the ethane, ethene and

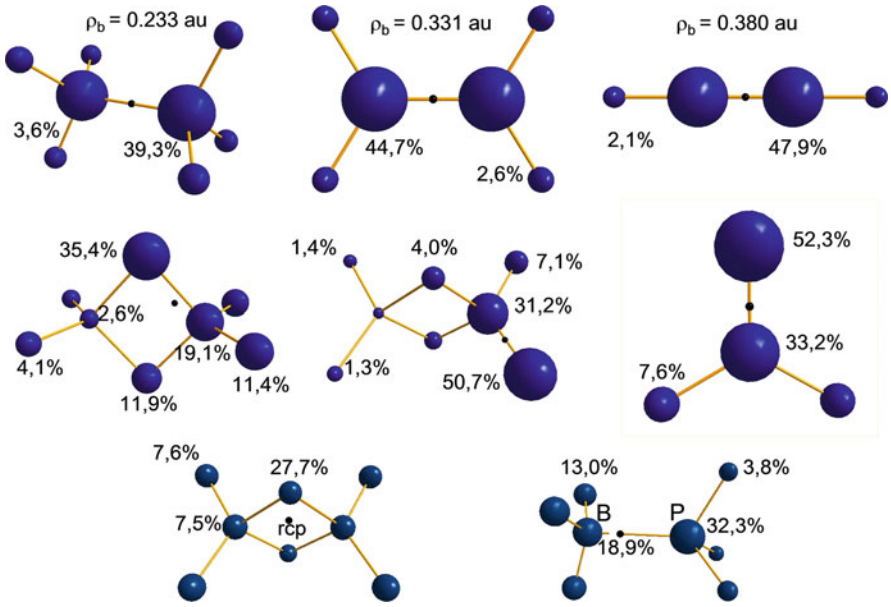


Fig. 2 Percentage source function contributions. *Top* (*from left to right*): ethane, ethene, ethyne, rp: C–C bcp. *Middle* (*from left to right*): B_2H_6 (rp: B–H_{bridge} bcp), B_2H_6 (rp: B–$H_{terminal}$ bcp), BH_3 (rp: B–H bcp); *Bottom* (*from left to right*): B_2H_6 (rp: rcp) and BH_3–PH_3. The rp position is indicated by a *black dot*. Each atom Ω is displayed as a *sphere* with volume proportional to the SF percentage contribution from Ω to the electron density at the shown rp (adapted from Fig. 3 with permission from [14]. http://pubs.rsc.org/en/Content/ArticleLanding/2007/FD/b605404h. Copyright 2007, The Royal Society of Chemistry (RSC) and from Figs. 4 and 5 with permission from [33])

Table 3 Source function description of C–C bonds in CH_n–CH_n (n = 1–3) hydrocarbons[a,b]

System	N(C)	R_e, Å	ρ_b	S($\mathbf{r}_b$,C)	S($\mathbf{r}_b$,H)	S%($\mathbf{r}_b$,C+C′)
Ethane	6.013	1.542	0.233	0.091	0.008	78.6
Ethene	6.075	1.349	0.331	0.148	0.009	89.4
Ethyne	6.159	1.225	0.380[c]	0.182[c]	0.007	96.0[c]

[a]Some of the data from [14]; if not otherwise stated, all quantities in au; rp is the C–C bcp
[b]DFT/BP86 [34–36] D95V Dunning–Hay basis set [37]
[c]There is a nonnuclear attractor [23] at the C–C midpoint; the reported source from each C atom includes half the contribution from the central nonnuclear basin

ethyne series), or with not enough electrons for a Lewis structure to be written (the 3c–2e bridging bonds in diborane) or instead forming a dative bond in a typical Lewis acid–base adduct (BH_3–PH_3). Atoms in Fig. 2 are displayed as spheres with volume proportional to their percentage source function contribution, S%($\mathbf{r}_p$,Ω), at the selected bcp.

As expected for a well-known covalent bond, most of the SF contributions to the C–C bcp density come from the two neighboring C atoms, with percentage contributions increasing with increased formal bond order from one to three (78.6%, 89.4%, and 95.8%), increased bcp density and C atom population (Fig. 2; Table 3) [9, 14]. The collective percentage contribution from the H atoms accordingly decreases along the series and so also does the percentage contribution from a single H atom, being 3.6% in ethane, 2.6% in ethene, and only 2.1% in ethyne, although the H source remains almost constant and small in value through the series (Table 3). Since both the percentage contributions from C and the density at bcp increase, the source function contributions from C significantly boost along the series, being doubled in ethyne, S(C) = 0.182 au, with respect to ethane, S(C) = 0.091 au (Table 3). Both C and H SF contribution and percentage trends agree with the increased strength, *s* character, and localization of C–C bonds along the series.

Comparing the bridging and terminal B-H bonds in diborane, one immediately notices [14] important differences in the source contributions from the two linked atoms. They exceed 80% for the terminal B-H bond, analogously to the B–H bond in BH_3 (Fig. 2), while they are as low as 54.3% for the bridging bond, with the residual contribution, except for a fourth of it, being shared almost equally among the other H_{bridge} atom and each of the two terminal H atoms closest to the reference bcp. Note that for both bonds and contrary to the case of C-C bond in hydrocarbons, the contributions from the two bonded atoms is largely asymmetric, the more electronegative atom – the hydridic H – contributing in both cases over 60% of the bcp density *determined* by these two atoms.

Despite being involved in the so-called 3c–2e bond, the contribution to the density at the B–H_{bridge} bcp from the other B′ atom is less than 3%. A similar description arises, however, when the delocalization indices $\delta(\Omega,\Omega')$ [38] are analyzed, although delocalization indices and source function contributions are not physically related in a direct way [14]. Delocalization indices are obtained through double integration of the pair density $\pi(\mathbf{r}, \mathbf{r}')$ over the basins of atom Ω and Ω′, with the electrons being kept in separate basins and provide a quantitative

measure of the electron sharing between basins A and B. They convey information about the special electronic connection between two basins, whereas the $S(\mathbf{r}, \Omega)$ values simply relate the Laplacian electron distribution in one basin with its contribution to the electron density value at a single, given point. However, when this reference point is the bcp, taken as the most representative point for the interaction between two linked atoms, one may envisage that delocalization indices and source function contributions might empirically be related in some way and under special circumstances. For instance, when bond covalency is known to increase along a series of related compounds, one anticipates that both the delocalization index between the covalently bonded atoms and their source function contributions to the electron density at the associated bcp will correspondingly increase along the series. In the hydrocarbon series discussed earlier, we indeed found [14] that the increased SF percentage contributions from the C atoms is paralleled by an increased delocalization index, from 1.03 in ethane, to 1.92 in ethane and 2.76 in ethyne. Conversely, the very small source function percentage contribution from the C atom to the bcp density of the C′–H′ bond (1.7%, 4.0%, and 4.7% for the three members of the series) is paralleled by a corresponding δ(C, H′) value as low as 0.042, 0.062, and 0.084, in ethane, ethene, and ethyne, respectively. Coming back to the diborane case, the δ(B,B′) value is 0.084, while the electronic share between the B and the bridged hydrogen is about four times as large, $\delta(B,H_{bridge}) = 0.332$. Moreover, both the electron share between the two bridged hydrogens, $\delta(H_{bridge},H'_{bridge}) = 0.214$ and that between a terminal and a bridged H, $\delta(H_{bridge},H) = 0.109$, are (much) greater than the electron sharing between the two B atoms. All these results are in line with the delocalized nature of SF contributions to the density at the B–H_{bridge} bcp and the small SF contribution to such density from the other B atom. Conversely, the localized description of source contributions to the bcp density of the terminal B–H bond properly complies with a delocalization index value for this pair of bonded atoms, $\delta(B,H) = 0.59$, which is almost twice as big as that for the B and the bridged hydrogen atom. Figure 2 also shows the large delocalization of source contributions to the electron density at the diborane ring critical point.

A spread of sources is also observed in the case of the Lewis adduct, suggesting that its formation involves the two molecules in their entirety, rather than the electron-rich and the electron-deficient atoms only. Indeed, the BH_3 and PH_3 moieties are found to determine to similar extent the density at the bcp of the adduct, with even larger contribution from the acidic moiety, $S(BH_3) = 58\%$ [33]. However, despite being much closer to the bcp, the electron-deficient B atom contributes only about one half the source from the P atom and are the much larger sources due to its three linked hydridic H, as compared to those due to the acidic H linked to P, that eventually lead to the dominance of the acid moiety contributions.

3.2.2 Cyclic Conjugated and Aromatic Hydrocarbons

Table 4 lists source contributions to the C–C bonds in three cyclic conjugated and aromatic hydrocarbons (see Scheme 2 for atomic numbering). As anticipated earlier

Table 4 Source function description of C–C bonds in cyclic conjugated and aromatic hydrocarbons[a]

System	Bond C_i–C_j	R_e, Å	ρ_b	S%(C_i + C_j)[b,c]	S%(H_i + H_j)[b,c]	S%(C_{inn} + C_{jnn})[b,c]	S% ($\Sigma_{other\ C}$)	S% ($\Sigma_{other\ H}$)
Benzene	C_1–C_2	1.411	0.300	84.7 (*0.254*) (84.4, 0.3)[e]	5.7 (*0.017*) (7.3, −1.6)[e]	5.3 (*0.016*) (3.4, 1.9)[e]	1.4	2.8
		(0.5z)[d]	0.245	82.5 (*0.202*) (73.8, 8.7)[e]	6.6 (*0.016*) (8.5,−1.9)[e]	6.1 (*0.015*) (3.8, 2.3)[e]	1.6	3.4
		(1.0z)[d]	0.141	72.6 (*0.101*) (45.2, 27.4)[e]	10.2 (*0.014*) (13.4,−3.3)[e]	8.9 (*0.013*) (4.9,4.0)[e]	2.6	5.6
		(1.5z)[d]	0.065	50.6 (*0.033*) (3.1, 47.5)[e]	18.5 (*0.012*) (25.3,−6.8)[e]	14.8 (*0.010*) (6.5, 8.3)[e]	4.4	11.6
		(2.0z)[d]	0.027	4.0 (*0.001*) (−63.0, 67.0)[e]	35.4 (*0.010*) (50.7, −15.3)[e]	24.3 (*0.006*) (5.8, 18.5)[e]	9.4	27.1
Cyclohexadiene	C_1–C_2	1.358	0.328	87.2 (*0.286*)	5.3 (*0.017*)	3.2 (*0.010*)	0.8	3.7
	C_1–C_6	1.474	0.267	81.9 (*0.219*)	6.2 (*0.017*)	7.2 (*0.019*)	0.4	4.1
	C_2–C_3	1.518	0.247	79.6 (*0.196*)	10.3 (*0.025*)	4.9 (*0.012*)	1.2	4.1
	C_3–C_4	1.566	0.225	75.9 (*0.171*)	14.5 (*0.033*)	4.1 (*0.009*)	1.9	3.5
Cyclohexene	C_1–C_6	1.350	0.331	87.4 (*0.289*)	5.3 (*0.018*)	2.5 (*0.008*)	0.1	4.4
	C_1–C_2	1.508	0.251	79.6 (*0.200*)	10.2 (*0.025*)	5.0 (*0.013*)	0.1	5.0
	C_2–C_3	1.558	0.228	76.4 (*0.174*)	14.6 (*0.033*)	3.4 (*0.008*)	0.9	4.8
	C_3–C_4	1.568	0.224	76.1 (*0.170*)	15.1 (*0.033*)	2.4 (*0.005*)	1.0	5.2

[a]Atomic numbering in Scheme 2. If not otherwise stated, all quantities in au; DFT/BP86 [34–36] D95V Dunning–Hay basis set [37]

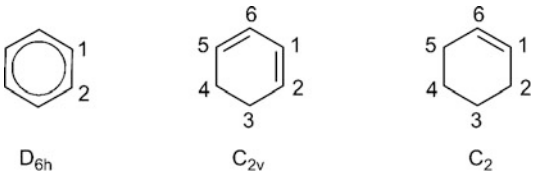
[b]SF value in parenthesis and in italic

[c]C_{inn} and C_{jnn} are the nearest neighbor C atoms of C_i and C_j, respectively. H_i and H_j are the H atoms linked to C_i and C_j, respectively

[d]Points along an axis perpendicular to the *xy* molecular plane and passing through the C–C bcp are taken as rps. The *symbol* denotes the *z* distance (au) of the rp from the molecular plane

[e]σ and π percentage contributions, in this order, with $\sigma\% + \pi\% = S\%(X_i + X_j)$, X = C or H

Scheme 2 Atomic numbering in benzene, cyclohexadiene, and cyclohexene



(cfr Sect. 2.3), the interesting question arises as of whether the SF tool might be able or not to mirror the model of π-conjugation in some way.

Part of the results listed in the Table 4 were presented by us at the Gordon Research Conference on Electron Distribution and Chemical Bonding in Oxford in 1998 and are reviewed and expanded here to challenge a recent too rigid assertion according to which "the π-electron delocalization in the benzene ring is not manifest in the SF when the rp is taken at the C–C bcp" [12]. The reasoning behind this statement was the null contribution from π molecular orbitals to the electron density in their nodal plane. However, as we discussed earlier in Sect. 2.3, σ- and π-distributions are not independent, but self-consistently interrelated [6, 20], and one may thus argue that some, perhaps small effect of electron conjugation be manifest also when the rp lies in the π-nodal plane, even though π-orbitals do not obviously yield *direct* contributions to the electron density in this plane. Results shown in Table 4 demonstrate that *this is actually the case*. Consider first the C–C bcp for the shortest bond(s) in each system, i.e., the one(s) with largest double-bond character. Both the SF and the S% contributions from the C atoms other than those directly involved in such a bond decrease with increasing double-bond character and electron localization. The S% contribution from the nearest neighbor C atoms decreases from 5.3% in benzene, down to 3.2% in cyclohexadiene, and to 2.5% in cyclohexene, while that from the two next-nearest neighbor C atoms is definitely smaller and decreases even faster with decreasing π-electron delocalization (1.4%, 0.8%, and 0.1%). Conversely, the contributions from the two atoms involved in the bond increase both in value and percentage through the series. All these trends nicely fit with a decreased π-electron delocalization and enhanced localized nature of the shortest C–C bond along the series. Also, the C1–C6 bond adjacent to the two double bonds in cyclohexadiene shows a very large SF contribution from the two nearest neighbor C atoms, C2 and C5, in agreement with the decreased bond length, 1.474 Å, with respect to standard single bond and partial double-bond character due to the π-electron delocalization mechanism. As shown in Table 4, SF and S% contributions from the nearest neighbor C atoms to the bcp density of the three unique C–C "single" bonds in cyclohexadiene decrease with increasing bond length distance and decreasing partial double-bond character, again in agreement with a parallel decreased importance of the π-electron resonance forms involving these bonds.

We move now one step further. We showed that the SF tool translates π-electron delocalization in enhanced S% contributions to the C–C bond bcp density from neighboring C atoms related to such a bond by electron delocalization mechanisms. Enhancement trends nicely follow an increased ability to realize π-electron

delocalization. One argues that such an increased S% contribution should become even more evident when analyzed using reference points for which the effect of π-electron conjugation takes place directly through π-electron distribution, rather than indirectly through σ–π electron interdependency. Table 4 indeed shows that the S% contributions to the C–C bond bcp density from the neighboring C atoms increase dramatically in benzene when the rp is moved along a line perpendicular to the bond path and directed above (or below) the molecular plane. The S% contribution from the nearest neighbor C atoms increases from 5.3% to 8.9% at 1 au above the molecular plane and becomes as large as 24.3% at 2 au above such plane. Contribution from the next-nearest neighbor C atoms is smaller, but also largely increases with increasing distance from the molecular plane, reaching a value of about 10% at 2 au above the plane.

Note that this ability of the SF to reflect π-electron conjugation is totally independent from a σ and π separation of the electron density since the SF tool has been applied to the total density. Were this separation not realizable, the same results shown in the Table 4 could have been obtained by analyzing an equivalent density distribution, albeit expressed in a completely different form, e.g., numerically or in terms of multipole model pseudoatom contributions. This observation is of great importance in view of the possibility to recover and quantify electron conjugation effects both when using electron densities derived experimentally (hence without σ and π separation being allowed) and when a departure from perfect symmetry would inhibit anyhow a proper separation of σ and π electron contributions. Such a use of the SF has not yet been explored in detail despite it appears to be very promising. Clearly the same would hold true for more complicate situations, like in organometallics, where a mixing of σ, π, and δ contributions can be envisaged for some bonds.

Since we deal here with densities derived from a molecular orbital approach, the separate, though interrelated, σ and π contributions to the source function values can also be precisely quantified. These are listed in Table 4, in the second-row entry for each of the considered distances from the molecular plane. The σ% and π% contributions from two bonded C atoms to the density at their bcp in benzene are shown to significantly decrease and, respectively, increase with distance from the molecular plane, as anticipated from the different relative weight of the associated electron distributions with such a distance. The corresponding SF contributions from the nearest neighbor C atoms show that the observed five times increase of their percentage total contribution to the density, on passing from the C–C bcp on the molecular plane (5.3%) to a point 2.0 au above this plane (24.3%) is the result of a less than a two times increase (from 3.4 to 5.8%) in the σ% contribution and of a dominant ten times increase in the π% contribution (from 1.9 to 18.5%). It is thus this latter contribution which leads to an increasingly importance of the nearest neighbor C atoms in determining the electron density of the benzene C–C bond when the rp for this bond is moved above or below the molecular plane. Note that the contributions from the nearest neighbor H atoms follow an opposite trend. Although their total S% contribution also increases with increasing distance from the molecular plane, the observed enhancement is the result of a seven time increase

of the σ% contribution, which largely exceeds the increasingly negative π% contribution. The opposite σ% and π% contributions trend for next neighbor C and H atoms is clearly a consequence of the presence in the former and lack in the latter of the *p* atomic orbitals and electrons.[6]

3.2.3 Second-Row Diatomic Hydrides

In Sect. 3.1, application of the SF to the Li–*X* (*X* = F, O, N, Cl, H) series was reviewed to show how such a tool is able to assess whether an almost perfect or only a compensatory chemical transferability of the Li atom – the fixed element through the series – characterizes the given set of related compounds. In the following, the SF description of the second-row diatomics H–*X* (*X* = Li, Be, B, H, C, N, O, F) is instead briefly reviewed [9] to inspect how this function is able to account for the well-known change of nature of the H atom – the fixed element of this series – with the change in the electronegativity of *X*. Figure 3 shows contour maps of

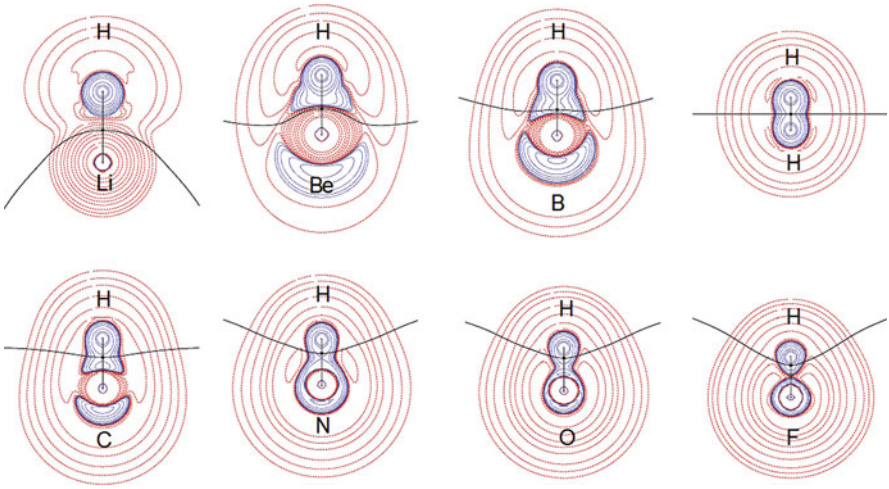


Fig. 3 H–*X* diatomics: contour maps of $L(\mathbf{r}) = -\nabla^2\rho$ *dashed red contours*, $L(\mathbf{r}) < 0$, indicate regions of charge depletion and *solid blue contours*, $L(\mathbf{r}) > 0$, denote regions of charge concentration. The bond path and the intersection of the interatomic surface with the plane of the map are shown for each diagram. The H-basin shape and the position of the bcp reflect the transition through the series from closed-shell to shared atomic interactions and from a cationic to anionic nature of the hydrogen. Changes in the H-basin shape and in its $L(\mathbf{r})$ distribution significantly affect both $S(\mathbf{r}_b,H)$ and the percentage contribution of the H atom to the bcp electron density (see Table 5) (adapted from Fig. 1 with permission from [9], Copyright 2003, Wiley-VCH Verlag GmbH & Co, KGaA)

[6]Although H atoms do not possess p-electrons, H basins as defined by QTAIM, may nonetheless contain π-electron MO contributions to their electron density.

$L(\mathbf{r}) = -\nabla^2\rho$ and the intersections of the interatomic Li|X surface with the plan of the map for all investigated diatomics. Changes in the H-basin shape and in its $L(\mathbf{r})$ distribution are anticipated to have a significant effect on both the SF value and its percentage contribution to the bcp electron density. The shape of the H-basin and the position of the bcp reflect the transition through the series from closed-shell to shared atomic interactions and from a cationic to an ionic nature of the H atom. On going from H–Li to H–F, the bcp moves from a region closer to the X nucleus of negative L, hence positive Laplacian, to a region closer to the H nucleus and with increasingly smaller Laplacian values (this is detailed in Table 5, where the symbol R_H denotes the distance from the H nucleus to the bcp). As a consequence, the S($\mathbf{r}_b$,H) value is found to increase from 0.02 au in HLi up to about 0.140 au in HO and HF. However, the ρ_b value increases along this series by a quantity about three times as large, namely from 0.033 au to 0.354 au, leading to a monotonic decrease in the S%($\mathbf{r}_b$, H) values along the series from about 60% in HLi to 38.8% in HF. Therefore, the decreasing electron population of H, from 1.874 e^- for the hydridic H in HLi to the only 0.399 e^- for the almost cationic H in HF and the corresponding increase of X's electronegativity, translates into a significantly decreased H's source function percentage contribution to the bcp density through the series. Attention has been called [9] to the fact that the decreasing percentage share from H atom with increasing electronegativity of X complies with the previously described parallel decrease of the local Shannon entropy in the H-basin, taken as an indication of an increasingly structured and localized H-atom's density along the series [39].

The LS, LG, and LV profiles along the bond path have been analyzed in detail for the H–X series [13], with the bcp being taken as rp. Only the main conclusions of that study by Gatti and Bertini are reviewed here. The LS profiles were found to show an increasing asymmetry, with respect to the bcp location, as the X's electronegativity increases through the series. The observed behavior was proposed to set up a clear and sensitive indicator of the increased (decreased) asymmetry and polar character of a bond along a series or of the changes of these features when the bond is placed in different chemical environments. From HC to HF, the LS asymmetry is the result of a corresponding asymmetry of LV, the term related to the potential

Table 5 Bond critical point properties and source contributions to its density ρ_b in the second-row diatomic hydrides[a]

H–X	R_e	R_H%[b]	q(H)[b]	ρ_b	$\nabla^2\rho_b$	S%($\mathbf{r}_b$,H)
H–Li ($^1\Sigma^+$)	3.016	54,1	−0.874	0.033	0.155	60.1
H–Be ($^2\Sigma^+$)	2.538	56,4	−0.798	0.083	0.188	60.3
H–B ($^1\Sigma^+$)	2.328	53,1	−0.574	0.163	−0.488	52.2
H–H ($^1\Sigma_g^+$)	1.400	50,0	0	0.243	−1.012	50.0
H–C ($^2\Pi$)	2.116	37,6	−0.052	0.245	−0.664	44.8
H–N ($^3\Sigma^-$)	1.958	30,2	0.219	0.305	−1.161	43.2
H–O ($^2\Pi$)	1.833	23,7	0.450	0.340	−1.548	41.0
H–F ($^1\Sigma^+$)	1.733	20,2	0.601	0.354	−1.686	38.8

[a]Data from [9], all quantities in au

[b]R_H% is the distance from the H nucleus to the bcp, expressed as bond length percentage; q(H) is the QTAIM net charge of H atom

energy, which was found to dominate LS for these systems around the bcp. Instead, the opposite holds true for HLi and HBe molecules where it is the sink contribution related to the kinetic energy, LG, which largely imparts a form to their LS profiles close to the bcp. In shared interactions, where electronic charge is accumulated along the bond, the LV term dominates in regions around the bcp and so determines positive source contributions at this point. On the other hand, small and negative local source contributions from these same regions are expected in the case of the nonshared interactions (HLi, HBe), where the electronic charge is removed from the internuclear space and separately accumulated (or transferred) within the two interacting basins, and the LG term prevails. Needless to say, this result complies with the standard picture of the $\nabla^2\rho_b$ classification of these extreme classes of chemical interactions. The LS, LG, and LV profiles serve as a magnification lens for the above-mentioned features and for those associated with an asymmetric sharing of density from the two interacting partners. Gatti and Bertini [13] also commented that the use of the LS, LG, and LV profiles adds to and detracts from the central role of the bcp in bond classification schemes, since the profiles use this point as a rp, but at the same time bring information on the effect on sources of all remaining points along the bond, thus avoiding the drawback of restricting the attention to the bcp only.

3.3 SF Description of Hydrogen-Bonded Systems

The hydrogen bond is with no doubt the most important intermolecular interaction [40]. It is ubiquitous in nature, from molecular aggregates in gas phase and in solutions, to inorganic materials and to biologically active macromolecules, playing a fundamental role in determining the stability, dynamical and reactivity properties of all these systems. The hydrogen bond (HB), which may be designated as a D–H···A interaction, shows an extraordinary variety of geometries and of dominating energetic contributions, according to the nature of the H-donor, D, and that of the H-acceptor atoms, A [40]. The HB energies are known to extend from 15 to 50 kcal/mol for the very short strong hydrogen bonds (SSHB), down to 1–4 kcal/mol for the weak bonds, i.e., from bond strengths typical of the covalent or ionic bonds bond to those characterizing the weak electrostatic or the van der Waals interactions. Relationships among the geometrical, energetic, electronic, and reactivity features of the various classes of HBs have been largely investigated by combining a variety of techniques [40], including, among other, structural determinations through X-ray and neutron diffraction, thermochemical measurements, infrared, RAMAN and NMR spectra, ab initio computations, topological studies of the electron density [41–44], or using other more sophisticated descriptors [45] and approaches [46]. Given these premises, why should one introduce one more descriptor, the source function, to characterize this widely studied interaction? Using a simple tool which may be applied with no change to both experimental and theoretical densities and exploring whether the SF contributions are able to

reflect the diverse nature of HBs certainly represented two good motivations behind such a test [9]. Enhanced SF contributions to the HB bcp density from the atoms most directly involved in the H-bond (D–H···A) and a parallel decrease of those from the remaining atoms, with increasing energy, covalency, and local character of the HB, could perhaps be anticipated. However, the extremely varying role of the H atom involved in the HB with change in the HB nature [9], or the existence of specific signatures for the low-barrier hydrogen bonds (LBHB) [47] and for the π-bond cooperativity mechanism controlling resonance-assisted hydrogen bonds (RAHBs) [9, 48] were certainly less trivial and predictable outcomes.

In the following, we review the main results derived from the only five HB studies [9, 13, 47, 49, 50] that, to the best of our knowledge, have thus far made use of the SF approach. We also add a number of new results, especially for the RAHBs systems, that come from closer examination of or from comparison among these studies. Not surprisingly, in view of the postulated relevant role of the LBHBs in enzymatic catalysis, the first paper due to Overgaard et al. [47] concerned the comparison of the SF contributions for two small molecules taken as examples of a low-barrier and of a single-well HB (benzoylacetone and nitromalonamide, respectively). Appreciation of the results of this pioneering study is facilitated by reviewing first the main outcome of a systematic study on a number of paradigmatic HB systems carried on by Gatti et al. [9].

Evaluation of the SF contributions to the density at the HB bcp along the reaction path for two water molecules, which approach each other within the linear dimer C_s constraint, served as a model to study the effect on the localizability or spread of atomic sources when the H···O and O···O distances change from the values typical of the weak isolated HBs to those characteristic of the charge-assisted H-bonds [48]. Gatti et al. [9] found that, despite an almost constant and comparable source function contribution from the donor and the acceptor water molecules, the atomic S% contributions vary radically along the bond path (Table 6). The S%(H) clearly

Table 6 Source contributions to the hydrogen-bond bcp density in a number of prototypical hydrogen-bonded complexes[a]

System[b]	$R_{O\cdots O}$, Å	$R_{H\cdots O}(R_{O-H})$, Å	$\nabla^2\rho_b$	S% (H)	S% (D)	S% (A)	S% (H+A)	S% (H+D+A)
1 +(CAHB)	2.409	1.204 (1.204)	−0.415	31.4	9.6	51.7	83.1	92.7
2 – (CAHB)	2.430	1.216 (1.214)	−0.392	32.1	8.3	49.9	82.0	90.3
4 (RAHB)*[c]	2.370	1.209 (1.209)	−0.425	32.2	8.5	48.8	80.9	89.5
3 (RAHB)	2.538	1.639 (1.008)	0.148	2.1	34.7	34.0	36.1	70.8
5 (PAHB)	2.749	1.850 (0.984)	0.092	−14.4	53.1	31.0	16.6	69.7
6 (IHB)	3.020	2.077 (0.943)	0.067	−72.3	106.6	18.7	−53.7	53.0
	2.750[d]	1.809 (0.941)	0.124	−35.5	74.2	32.2	−3.2	71.0
	2.500[d]	1.564 (0.936)	0.216	−12.5	55.4	41.1	28.6	83.9
	2.250[d]	1.327 (0.923)	0.333	+2.8	39.6	46.2	49.1	88.7
	2.000[d]	1.110 (0.890)	0.208	+13.1	29.6	50.8	63.8	93.5

[a]Data from [9]. If not otherwise stated, all quantities in au. D and A are the H-donor and H-acceptor oxygen atoms, while H is the hydrogen atom involved in the H-bond

[b]H-bonded complexes are numbered as in Fig. 4c and classified according to [48]

[c]Transition state for the H atom migration in malonaldeyde

[d]Points along the reaction path for the approach of two water molecules within the linear dimer C_s constraint

appeared as the most characteristic marker of the HB nature along the reaction path, being largely negative at equilibrium distance (O···O = 3.020 Å) and becoming progressively less negative and finally positive for very short O···O and H···O distances. Furthermore, the sum of the percentage contributions from the H and the acceptor O atoms to the density at their HB critical point bcp was found to change from a very negative value at equilibrium distance, to a value almost of zero at distances (O···O = 2.750 Å) typical of the long chains of O–H···O bonds in water and alcohols where σ-bond cooperativity effects take place [51], and to finally change to positive values, yet as large as only 50%, at the O···O = 2.250 Å distance typical of the charge-assisted H-bonds [48]. This result clearly contrasts the standard situation of covalent bonds where the sum of S% contributions from the two bonded atoms to their bcp density is usually 80–90% or more. One gets comparable percentage sources only when the percentage contribution from the donor O atom is also included in this sum (Table 6) and for O···O distances below 2.5 Å, indicating that even at such (short) distances the HB retains at least a three-center nature [40, 48]. At equilibrium distance, atoms other than the three atoms more directly involved in the H-bond determine almost half of the density at the HB critical point, a result that clearly provides a fairly delocalized picture of sources for such a bond at equilibrium. The source spread has been related [9] to the dominant electrostatic nature of the H-bond at these distances, to be contrasted with the partial covalent character of the bond when the S(H+A+D)% value becomes dominant and the atomic sources much more localized.[7]

Figure 4a, b display the percentage atomic sources for three representative points along the HO–H···O–H_2 reaction path and the corresponding contour maps of the negative Laplacian $L(\mathbf{r})$ in the symmetry plane containing the donor water molecule and the HB bond path.

The dramatic change of sources with changes in the H-donor to H-acceptor distance is clearly evident from the picture. Inspection of the $L(\mathbf{r})$ contour maps (Fig. 4b) around the H atom involved in the HB reveals why the S% contribution from this atom changes from highly negative to moderately positive along the bond path [9]. At equilibrium distance, where S%(H) is as negative as −72.3%, the H atom exhibits a truly asymmetric shape and a clearly inhomogeneous $L(\mathbf{r})$ distribution within its basin. A small region of charge concentration includes and surrounds the O–H bcp, yielding a standard positive S(H) contribution to this point, while the HB bcp lies in a much larger region of positive Laplacian, so providing negative

[7]Delocalization of sources should not automatically be always associated with an electrostatic interaction. It may clearly also occurs for weak van der Waals interactions or for the source contributions to the density at a rcp of a *n*-membered ring, with *n* greater than 4 and at a molecular geometry far from a structural catastrophe point. When more than the two bonded atoms are present, localization of sources is instead certainly related to a (partial) covalent character for the investigated bond; π-conjugation, as discussed earlier in this chapter, leads to an enhanced site-specific delocalization of sources with respect to a corresponding case where it is not operating.

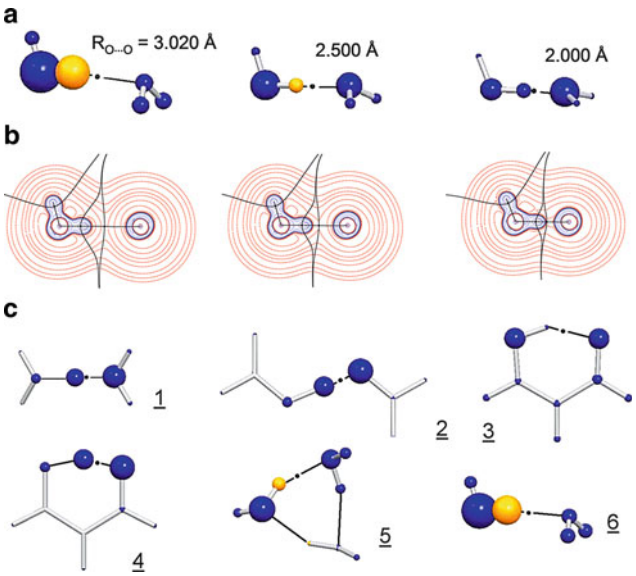


Fig. 4 Source function in hydrogen-bonded systems. (**a**) SF atomic contributions to the electron density at the hydrogen bond bcp (*denoted by a dot*) as a function of the donor to acceptor oxygen atom distance along the reaction path for the approach of two water molecules. Atoms are displayed as *spheres* whose volume is proportional to the SF percentage contribution from Ω to the electron density at bcp. Positive (negative) sources in *blue* (*yellow*). (**b**) Portraits of the negative Laplacian, $L(\mathbf{r})$, in a plane containing the H-donor molecule and the O···O internuclear axis for the corresponding systems reported in the first row. (**c**) SF atomic contributions to the electron density at the hydrogen bond bcp in a series of prototypical hydrogen-bonded complexes (identified by the same numerical labels listed in Table 6) (adapted from Figs. 5–7 with permission from [9]. Copyright 2003, Wiley-VCH Verlag GmbH & Co, KGaA)

local source contributions to the HB bcp density.[8] Because of the larger size and closer proximity of this region to the HB bcp, the negative local source contributions largely override the positive ones coming from the farther and smaller region of charge concentration. Conversely, as the O···O distance diminishes and the O–H and H···O distances become more similar one to the other, the shape of the H atom and the $L(\mathbf{r})$ distribution within its basin become both more and more symmetric, with the region of charge concentration and that of charge depletion, respectively, increasing and decreasing in their relative sizes. These changes clearly explain why the S%(H) contribution to the HB bcp density turns out to be less and less negative with decreasing donor to acceptor oxygen separation and becomes eventually

[8]Note that regardless of the relative size of the negative and positive $L(r)$ regions, the integrated $L(H)$ value should vanish because of the zero-flux QTAIM condition. The region of charge concentration is smaller because it is on average more concentrated than the region of charge depletion is on average diluted.

positive at the shortest displayed O···O distance. In brief, the complex variation occurring to shape, size, and Laplacian distribution of the H atom along the reaction coordinate $R_{O\cdots O}$ is nicely summarized by and translated into the change of a single-valued function of $R_{O\cdots O}$.

When the SF tool was applied to a series of prototypical HB complexes, most of the qualitative conclusions from the model study on water dimer were retained, but important new and subtle facets also emerged [9]. The investigated complexes are shown in Fig. 4c and classified according to Gilli and Gilli [48] as charge-assisted hydrogen bonds, [+CAHB, **1**: $(H_2O\cdots H\cdot\cdot OH_2)^+$; –CAHB, **2**: the open form of the formic acid-formate anion complex], resonance-assisted hydrogen bonds [RAHB, malonaldeyde, in its C_s equilibrium form, **3**, and in its C_{2v} transition state, **4**, for the H-atom transfer between the two oxygen atoms], polarization-assisted hydrogen bond [PAHB; **5**: cyclic homodromic[9] water trimer], and isolated hydrogen bond [IHB; **6**: water dimer at equilibrium geometry]. The SF percentage contributions listed in Table 6 and displayed in Fig. 4c neatly confirm S%(H) as a clear indicator of the H-bond nature. Its value shows even more pronounced variations with change in the donor to acceptor atom distance than found for corresponding distance changes along the approach of two water molecules. Enhanced changes with increasing or decreasing $R_{O\cdots O}$ are similarly found for the percentage sources from other atoms or group of atoms. For instance, S%(H) and S%(H+A) are already as large as about 32% and 82% for systems **1**, **2** at an $R_{O\cdots O}$ distance of about 2.4 Å, while they amount to about only 3% and 49% in the water dimer at the much shorter 2.25Å $R_{O\cdots O}$ distance. The percentage source from the two H-bonded atoms, S%(H+A), clearly denotes a marked covalent character for this bond in the ±CAHB systems, while it is not so for the neutral water dimer system when compressed at similar or even shorter donor to acceptor atom separations. At variance with the compressed water dimer, ±CAHB complexes exhibit equal or almost equal O–H and H···O distances, and it is this symmetric location of the H involved in the HB that leads to the enhanced percentage source contribution from H and to the pronounced covalent character of their HBs.

Analysis of the SF contributions for the RAHB in malonaldeyde, **3**, revealed quite an interesting fact [9]. The sum of percentage SF contributions from the triad of atoms directly involved in the RAHB did not follow the trend one would have expected from the HBs in the other members of the series, but rather showed a value, S% (H+D+A) = 70.8, which is comparable to that found in the cyclic water trimer, **5**, or in the water dimer at $R_{O\cdots O}$ 2.750 Å, despite $R_{O\cdots O}$ being in **3** shorter by 0.250 Å. A similar S%(H+D+A) contribution decrease, with respect to the expected trend, was also observed for the RAHB in the malonaldeyde TS, **4**. Its S %(H+D+A) value was found to be comparable and even slightly smaller than that for CAHBs **1** and **2**, despite their larger O···O separations.

[9]A cyclic $(H_2O)_n$ water polymer is called sequential or homodromic if each O atom is acting contemporarily as an H-donor and as an H-acceptor.

RAHBs are known [40, 48] to be characterized by the mechanism of π-bond cooperativity. It relies on the positive synergism between HB strengthening and π-delocalization enhancement that occurs when the H-donor and the H-acceptor atoms are connected by a short chain of conjugated single and double bonds. Such a π-bond cooperativity is revealed in **3** by a particularly large source function percentage contribution, S%(C) = 8.3, from the carbon atom linked by a conjugated formal double bond to the acceptor oxygen atom. It is this large S% value, also clearly evident in Fig. 4c, the main cause of the decreased percentage SF contribution from the triad of atoms directly involved in the H-bond. The source contributions from the C atom linked to the oxygen donor is, instead, less than half as large, S%(C) = 3.8, despite the two C atoms are almost equally distant (3.3 and 3.8 Å, respectively) from the HB bcp. The different source function contributions from these two carbon atoms seem to clearly reflect the diverse role these two atoms have in the π-delocalization mechanism leading to the tautomer of **3** through TS **4**. The C atom with greater S%(C) value releases its shared π-electron pair to form the O–H bond in the tautomer of **3**, while the other carbon, with an halved S%(C) value, can contribute to this electron pair shift only in an indirect way through π-electron conjugation. Although the HB bcp lies in the nodal plane of the π distribution, this distribution has an indirect effect on the σ density, hence on the charge density at the HB bcp, as already documented for the conjugated hydrocarbon systems in this review (cf. Sect. 3.2.2). Also in this case, the analysis of SF contributions along a line perpendicular to the molecular plane and passing through the HB bcp could magnify the enhancement of delocalization of sources, which is induced by the π-electron delocalization mechanisms leading to RAHBs.

Based on the results obtained from the investigated series of hydrogen-bonded prototypical complexes, Gatti et al. [9] proposed the following classification of OH···O interactions, in terms of the SF tool: (a) IHBs have typically a highly negative SF contribution from H, S%(H) < -70, highly positive S%(D) value, much larger than S%(A), and SF percentage contributions as big as about 50% from atoms other than the D,H,A triad, in agreement with a dominant electrostatic nature for such bonds; (b) PAHBs still exhibit a negative S%(H) value, but almost halved with respect to IHBs, and are characterized by a S%(D) value greater than the S%(A) value, but less than twice as big, and by a substantially smaller external percentage source contribution than found for IHBs; (c) RAHBs are distinguished by a very small, but positive source from H, comparable sources from the donor and acceptor oxygen atoms, and in particular by a SF contribution from atoms other than the D + H + A triad that, being similar to that found for PAHBs despite the significantly smaller O···O separation in RAHBs, nicely agrees with the π-electron delocalization mechanisms proposed [52, 53] for such peculiar class of HBs; (d) ±CAHB show large and positive S%(H) values (about 30%), an SF contribution from the acceptor oxygen atom which is about five times as large as that from the donor oxygen atom, and extremely high SF contributions from the two atoms linked in the H-bond, S(A+H) >80%, consistently with the localized and largely covalent nature of the H···A interaction in these charged complexes. The correspondence between this HB classification and the one proposed [45] using the ELF [54–56]

topological approach was also highlighted. The reader is addressed to the original paper [9] for details.

Gatti et al. [9] also examined HB systems where either the D or the A atom, or both of them are no longer O atoms. Although based on a limited number of cases, it was shown that the relative weights of the SF contributions from atoms in the HB complexes are related not only to the D and A atoms distances but also to the specific nature of the H-donor and H-acceptor atoms. The conclusion that "the source function seems sensitive enough to discriminate between different donor–acceptor pairs, despite similar donor to acceptor distances and that it constitutes a very suitable tool to get insights into the hetero-nuclear H-bonds" [9] is likely to be a valid one, but certainly would warrant a more systematic investigation. A fortiori this holds true for more subtle aspects like the discussed capability of the SF tool to disclose the signature of π-bond mechanisms leading to RAHBs, for which a comparative study of systems showing π-bond cooperativity or anticooperativity effects [57] could be definitely more compelling. Equally important would be exploring in some detail whether also the mechanisms of σ-bond cooperativity can be revealed by the SF tool, as the noticeable source contribution to the H-bond density from an hydrogen atom involved in a different HB of the homodromic water trimer would seemingly suggest (Fig. 4c).

We are now ready to review those results of the paper by Overgaard et al. [47] that are of relevance for the application of the SF tool to HBs. In their effort to provide an answer to the very important question of what causes the formation of an LBHB, the authors investigated three short NH···O HBs in a cocrystallized complex of betaine, imidazole, and picric acid, which serves as a model for the active site (the catalytic triad) in the serine protease class of enzymes. The occurrence of an LBHB had in the past been postulated as a condition for TS stabilization in the catalytic triad, which is in turn necessary to increase the proteases rate constant [47]. As anticipated earlier, Overgaard et al. also studied the OH···O bond in benzoylacetone (bza) and nitromalonamide (nma) as simple examples of an LBHB and of a single-well HB, respectively. This spectrum of HBs thus covered the complete range of strong HBs, that is "localized"[10] HB (the H atom was indeed found to be firmly localized in the "nitrogen wells" in all the three NH···O bonds of the catalytic triad), LBHB (bza), and single-well HB (nma). Clearly, it is not simply the donor–acceptor distance that determines the HB type, since, for instance, the authors noticed that citrinin has a localized HB, while bza exhibits a LBHB, notwithstanding their almost identical O···O separation. Moreover, based on the H···O distance and the topological properties at bcps, one HB in the catalytic triad complex was proved to be stronger and more covalent than the other two NH···O bonds, despite its largest O···O distance. The authors thus speculate that there must be differences in the chemical environment that change the potential energy surface

[10]Note that this definition of localized HBs refers to the shape of the DH···A potential well and to the consequent localization of the H close to the D atom, rather than to a "localization" of sources and related H···A covalency as described earlier for the strongest prototypical HBs.

Table 7 Geometrical data and source contributions to the hydrogen-bond (HB) bcp density in a number of systems with strong HBs[a]

System	HB[b]	$R_{H\cdots A}(R_{D-H})$, Å	$R_{D\cdots A}$, Å	ρ_b	S%(D)	S%(D+A)
Betaine, imidazole	HB1 = N1A–H1A···O1A	1.630 (1.046)	2.614	0.051	1.1	57.9
and picric acid	HB2 = N3A–H3A···O8	1.635 (1.057)	2.684	0.054	4.9	56.6
complex [47]	HB3 = N1B–H1B···O1B	1.681 (1.048)	2.676	0.046	−2.3	57.3
nma[c]	O1–H(*X*)···O3	1.308 (1.140)	2.391	0.122	23.4	61.1
bza[c]	O2–H(*X*1)···O1	1.329 (1.245)	2.502	0.129	30.4	55.5

[a]Data from [47]. If not otherwise stated, all quantities in au. D and A are the H-donor and H-acceptor, while H is the hydrogen atom involved in the H-bond
[b]Labeling of atoms involved in the HBs according to Fig. 5
[c]nma and bza are nitromalonaldeyde and benzoylacetone, respectively

of the hydrogen atom, and that such a surrounding effect on the HB character could be possibly traced out using the SF tool. Table 7 reports geometrical data for the five strong HBs investigated and compares their S% atomic contributions to the electron density at the HB bcp, while Fig. 5 displays the $L(\mathbf{r})$ portraits leading to such contributions.

The LBHB in bza, having an HB distance only slightly larger, 1.329 Å, than that of the very elongated O–H bond, 1.245 Å, exhibits a quite symmetric $L(\mathbf{r})$ distribution around the H atom, which translates in an impressively large SF contribution from H to the HB bcp density. Its value is 38% larger than in nma and represents 30.4% of the bcp density, to be compared with a corresponding value of only 23.4% for nma, despite this latter molecule exhibits much shorter $R_{O\ldots O}$ and slightly shorter HB distances (Table 7).[11] The LBHB is thus characterized by an enormously increased H atom source contribution, while the contributions from the atoms not directly involved in the HB are almost similar in the two systems (14.1% and 15.5% in bza and nma, respectively). The diminished percentage contribution from the donor–acceptor atom [S%(D) = 9% and 17% in bza and nma] compensates for the higher S%(H) contribution in bza due to the presence of an LBHB.

The strong, but localized HBs in the catalytic triad complex show, instead, the usual asymmetric $L(\mathrm{r})$ distribution around the H atom, with a shared region of negative Laplacian between the donor N atom and its covalently bonded H atom. As a consequence, the source percentage contribution from H at the HB bcp density is small, close to zero and even negative for HB3 (Table 7). HB energy wells leading to strong localized HBs are thus characterized by a much smaller source from H than for single-well or low barrier HBs and by a more than doubled, about 40%, SF percentage contribution from the surroundings (atoms other than D, A, or H). Despite its largest N···O separation, 2.684 Å, HB2 was found to have an $L(\mathrm{r})$ distribution slightly more polarized toward the acceptor atom and the largest source contribution from H, both in magnitude and percentage value, in agreement with

[11]Note also that S%(H) in bza is dramatically larger than in malonaldeyde **3** ($R_{O\cdots O}$ = 2.538 Å) and in water dimer at $R_{O\cdots O}$ = 2.5 Å, despite a quite similar donor to acceptor separation, $R_{O\cdots O}$ (bza) = 2.502 Å.

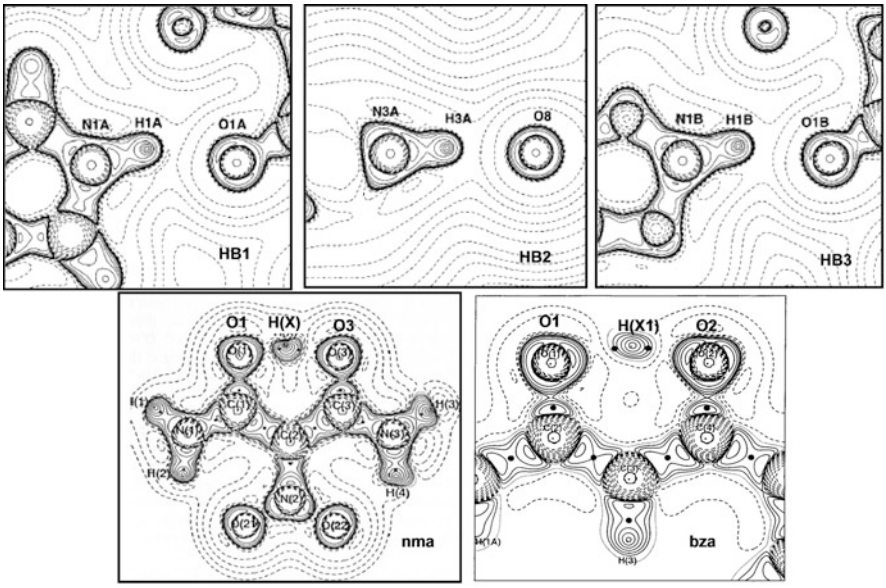


Fig. 5 Portraits of the negative Laplacian, $L(\mathbf{r})$, in a series of HBs covering the complete range of strong HBs: localized HB (*top*: HB1, HB2, and HB3 in the cocrystallized complex of betaine, imidazole, and picric acid), single-well HB (*bottom left* nitromalonaldeyde, nma), and Low Barrier HBs (*bottom right* benzoylacetone, bza). The different $L(\mathbf{r})$ portraits around the H atom involved in the HB lead to different S(H) and S%(H) values for the diverse HB types (see Table 7). The $L(\mathbf{r})$ portraits are obtained from ab initio computations, analogously to the source function contributions data reported in Table 7 (adapted from Fig. 4.1.4 of [58] and Fig. 4 with permission from [47], Copyright 2001, Wiley-VCH Verlag GmbH & Co, KGaA)

the stronger and more covalent character assigned to this bond on the basis of the bcp topological properties. Overgaard et al. [47] could thus conclude that HB2 can be characterized as an intermediate between localized HBs and delocalized LBHBs. The merit of their study was to show that the nature and strength of an HB are not "unambiguously determined by the geometrical parameters which define the system" [47], and that the formation of an LBHB is revealed by a dramatic increase in the H atom source contribution to the HB bcp density, whereas the changes in the contributions from the surrounding seem to be more subtle. While the SF neatly summarizes the changes in the $L(r)$ distribution around the H atom due to the formation of an LBHB, it yet appears not to be able to reveal the causes leading to the onset of an LBHB potential well. Further studies in this direction are surely required.

Another more recent study by Sörensen et al. [49] concerned the environmental influence on the electronic character of the HBs in a β-diketone, 2-acetyl-1,

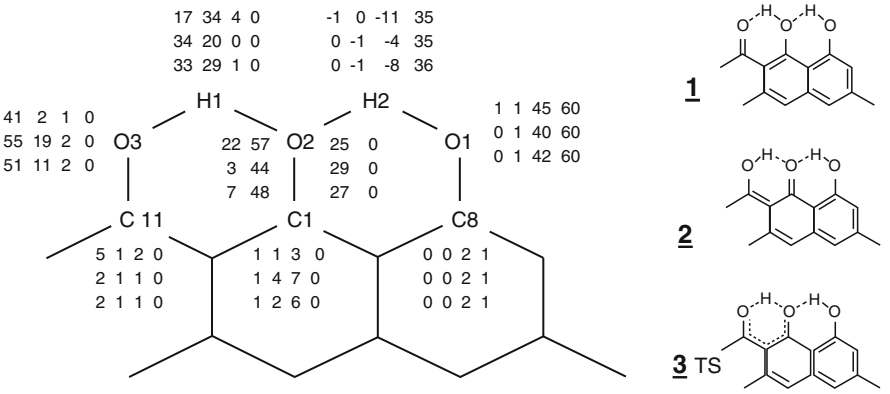


Fig. 6 Source function percentage contributions from the eight labeled atoms to the four O···H bonds of structure **1**, its quinone-like tautomer **2**, and their interconnecting TS, **3**. Sources are listed in matrices with rows corresponding to structures **1**, **2**, and **3** and columns 1–4 corresponding to the four O···H bonds, *from left to right* (O3···H1, H1···O2, O2···H2, H2···O1). O···H bond distances are listed for each structure in Table 8. Sources from atoms contributing less than 1% are not shown (adapted from Chart 1, with permission from [49], Copyright 2006, American Chemical Society)

8-dihydroxy-3,6-dimethylnaphtalene, **1** (Fig. 6), studied by low-temperature synchrotron X-ray diffraction and DFT calculations. Two short intramolecular O–H···O interactions are found in **1**, with one O···O distance being extremely short, 2.393 Å, similar to that found in nma, and the other being 2.598 Å short. Due to system's neutrality, both interactions should be termed as RAHBs, though the very short one lies outside the range typically associated with this class of HBs. However, at variance with nma, which has a perfectly symmetric keto–enol fragment despite a highly nonsymmetric HB, the keto–enol fragment in **1** is *far from being symmetric* and represents a severe outlier relative to published RAHB correlations [48, 52]. Furthermore, the two OH···O HBs were found structurally quite different, though being part of rings formed by the same kind and sequence of atoms. Investigation of system **1** thus served two very interesting purposes. On one hand, to shed light on why a very short O···O distance combined with apparent π-localization may realize, in contrast to the widely acknowledged belief that very strong HBs in neutral systems require to be assisted by resonance. On the other hand, to challenge the question of which chemical environment forces the neat distinction between the two HBs in the molecule. The SF tool was adopted, along with other investigative tools, to address these two issues.

To fully account for the experimentally observed structure and to also model the effect on HBs features of alternative π-bonding structures, DFT calculations were performed on molecule **1**, its quinone-like tautomer **2**, lying about 1 kcal/mole higher in energy, and their interconnecting TS **3**. (Fig. 6). Table 8 lists the bond

Table 8 Geometrical data and O···H bcp density in 2-acetyl-1,8-dihydroxy-3,6-dimethylnaphtalene, 1, its diphenolic tautomer 2 and their interconnecting TS, 3[a]

Bond	**1**, R_e (ρ_b)	**2**, R_e (ρ_b)	**3**, TS ; R_e (ρ_b)
O3···H1	1.419 (0.098)	1.065 (0.264)	1.153 (0.205)
H1···O2	1.040 (0.282)	1.366 (0.112)	1.229 (0.165)
O2···H2	1.749 (0.040)	1.689 (0.048)	1.715 (0.044)
H2···O1	0.977 (0.349)	0.988 (0.337)	0.984 (0.341)
O3···O2	2.393	2.372	2.333
O2···O1	2.598	2.573	2.588

[a]Data from [49]. See Fig. 6 for molecular structure drawings. Distances in Å, bcp densities in atomic units (au). Ab initio theoretical data

distances and bcp density values of the four O–H, or O···H interactions shown in the Fig. 6, which also displays the corresponding percentage SF atomic contributions from atoms contributing more than 1%. The SF analysis clearly distinguishes O2···H2 bond from the remaining O···H or O–H interactions, in all structures **1**–**3**. The sum of SF percentage contributions from O2, H2, and O1 atoms to the O2···H2 bcp density is only 59%, 65% and 61% in **1**, **2**, **3**, while the corresponding contributions for the other three bonds in all the three structures always exceed 80% of their bcp density. These results classify O1–H2···O2 as a short normal electrostatic HB and the others H···O bonds as short HBs with large covalent character (O1–H2 in **1**–**3** is a normally covalent O–H bond). Hence, despite the similarities of the six-membered rings to which they belong, H1 and H2 exhibit completely different S%(H) values. The former atom contributes largely to the bcp density of both bonds with its linked O atoms, while H2 shows a high and negative SF percentage contribution to the bcp density of the O2···H2 bond, typical of an electrostatic HB. The positive and large S%(H) values for H1 in **1** (17% and 34% for O3···H1 and H1–O2, respectively) denotes a partially symmetric $L(\mathbf{r})$ distribution for such an atom associable with a low-barrier or single-well HB. Indeed, high-level DFT calculations on structures **1**–**3**, including zero-point vibrational corrections, suggested the O3···H1 hydrogen bond to be a double-well potential, with a very low barrier between the two minima, so that in practice the H1 sits in a single-well potential. Sörensen et al. [49] also noticed the large differences in the contributions from O2, the only O atom common to the two rings, to the bcp densities of the distant O3–H1 and O1–H2 bonds. In none of the three structures does O2 contribute to the normal O1–H2 interaction, while it contributes significantly to the O3–H1 interaction in all three structures, and in particular in **1**, in which it provides 22%. These striking differences in S%(O2) denote electron delocalization only in the left keto–enol fragment and may be easily rationalized in terms of the HB classification based on the SF tool given by Gatti et al. (see earlier). The π-delocalization is, however, only partial since this fragment is far from being symmetric, as stated earlier. In order to retain the fully delocalized naphthalene structure, C11–O3 has to be mostly of double-bond character and C1–O2 must be a single bond, which is closely realized in **1**. Comparison of the S% contributions from the C atoms linked by a formal double bond to the H-acceptor

O atoms in **1** and in malonaldeyde (Fig. 4) reveals how the SF tool reflects the departure of **1** from a real RAHB system. As discussed earlier, the contribution from the carbonyl C atom in malonaldeyde is quite significant, S%(C) = 8.3, while that of C11 in **1** is only 5.0%. Analogously, the contributions from the C linked to the O donor atom are as big as 3.8% in malonaldeyde and only 1.0% for **1**.

Clearly the SF analysis alone is not able to provide an answer to what causes a strong O3–H1 bond in **1** despite the partial lack of the usual RAHB mechanisms.[12] Nor it may offer alone an answer to the occurrence of two quite different O···H interactions in the two six-membered rings of **1**. Yet, despite these limitations, it sheds important light on and quantifies the different involvement of the various atoms in determining the densities at the two HB bcps. One may assume these differences be related to the environmental influence on the HB geometry and strength.

The very recent study by Schmidtmann et al. [50] on polymorphic isocotinamide–oxalic acid molecular complexes with strong O···H···N hydrogen bonds has the great merit of applying, for the first time for HBs, the SF tool to both experimental and theoretical Laplacian distributions. To avoid inconsistencies between the two distributions due to the shortcomings of the multipole model, the theoretical structure factors obtained from the periodic calculation were refined using the same multipole model adopted for the experimental data. SF calculations were then performed with the XD2006 package [59], taking as input the parameters refined from either the experimental or the theoretical structure factors. Figure 7 displays experimental and theoretical $L(\mathbf{r})$ plots for the O···H···N SSHB between

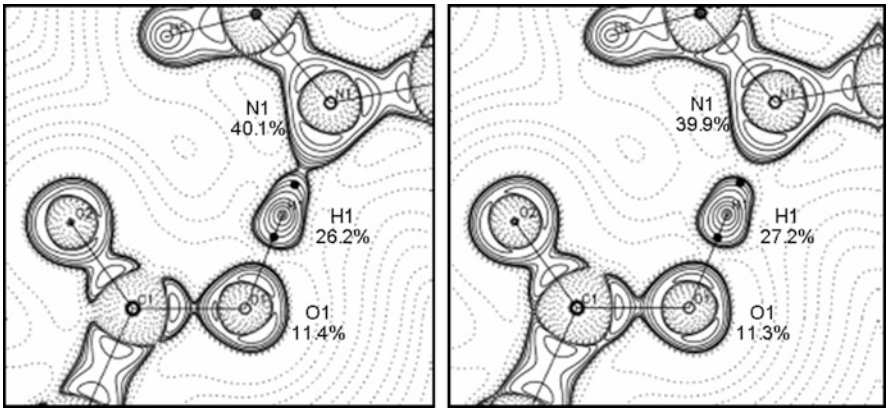


Fig. 7 Plots of the experimental (*left*) and theoretical (*right*) negative Laplacian $L(\mathbf{r})$ in the C1–O1–N1 planes of the O1···H1···N1 SSHB in the stable polymorph of the isonicotinamide–oxalic acid molecular complex. Source function percentage contributions to the density at the H1···N1 bcp are reported for the O1, H1, and N1 atoms (adapted from Figs. 4 and 5, with permission from [50], Copyright 2009, American Chemical Society)

[12]The decrease of the carbonyl atom contribution with respect to malonaldeyde could be compensated for by a distributed contribution from the carbon atoms of the conjugated naphthalene ring.

the acid and the pyridine base in the energetically stable polymorph of the complex.

It also reports the S% contributions from the O, H, and N atoms to the H···N bcp density. Both experiment and theory describe this HB as strong and largely covalent, with a high percentage source function contribution from the H, S%(H1) = 26.2 and 27.2, respectively, and also a high combined source contribution from the atoms comprising the HB, S%(O1 + H1 + N1) = 77.7 and 78.4, respectively.

The almost perfect agreement of SF values between theory and experiment seems apparently contradicted by the qualitative difference one observes in the $L(\mathbf{r})$ plots. The experimental map shows a continuous shared region of positive $L(\mathbf{r})$ enclosing both the H and the N nuclei, while such a region, regarded by Schmidtmann et al. [50] as a strong evidence for the covalent character of the H···N bond, is no longer continuous in the theoretical map. One should note, however, that qualitative topological changes may often occur as a consequence of small quantitative changes in the distribution of a scalar [2, 6, 8, 60]. A careful examination clearly reveals how close to a separation in two disjoint regions is the continuous region of charge concentration encompassing the H and N nuclei in the experimental map. Hence, this feature is neither stable – it is absent in the theoretical map and close to disappear in the experimental one – nor necessary for assigning a partial covalent character to this SSHB, as testified by the SF analysis and other examined bond properties. We will discuss later in this review (Sect. 4) how the SF descriptor, being based on a basin average of $L(\mathbf{r})$, often reveals itself [15] as a more robust descriptor of bond features against changes in the way $L(\mathbf{r})$ is obtained, than prove to be the $L(\mathbf{r})$ distribution itself or other commonly used topological indices. We therefore partly disagree with the authors' conclusion that the most convincing sign they provide for the covalent stabilization of the H···N SSHB, rather than coming from the characteristics of the SF or from the topological properties of the charge density, derives perhaps from the visualization of the Laplacian, showing a continuous region of charge concentration in the H···N bonds. This is a discontinuous sign and analogously to other important topological indices such as the presence or not of a bond path linking a couple of atoms whenever alternative pair of atoms compete for such a path [8], may be largely instable in given circumstances. Instead, the SF tool provides a smoothed description enabling one to meaningfully compare the larger or smaller covalent character of a series of related HBs.

We conclude this section on HB systems by briefly mentioning the use of the local source (LS) profiles to get a position space insight into the effects of HB intermolecular interactions on charge density distributions [13]. Gatti et al. [13, 44] have shown repeatedly that *deformation densities* (crystal density minus IAM density) and *interaction densities* (crystal density minus superposition of molecular densities) reach absolute minimum values in regions close to the HB critical point, being instead much larger in regions closer to the H nucleus or to

the nuclei of the heavier atoms involved in the HB. It thus appears that the properties at the HB critical point and in the region close to it contain the least information on the electron density polarization due to intermolecular interaction (and also to molecular formation, if IAM density is taken as reference). This observation clearly raises serious doubts about the use of the bcp properties only when discussing intermolecular interactions in crystals. Changes in atomic properties and related molecular properties, e.g., the molecular dipole, may be much more significant than changes in the HBs bcp densities [61, 62]. The analysis of the differences in the local source contributions to the HB critical point density using the crystal density or one of the model densities introduced earlier (IAM and superposition of the molecular densities) enables one to single out those remote molecular regions which mostly contribute to determining the small density changes at the HB critical point. Gatti and Bertini [13] discuss such changes by examining the difference LS profiles along the juxtaposition of the D–H and H···A (D=N, A=O) bond paths in the urea crystal. Conversely, analysis of these same LS difference profiles, but referred to rps located in those regions showing the largest variations in the crystal $L(\mathbf{r})$ distribution with respect to the model densities, might reveal those molecular regions which contribute more in determining such maximal changes. The interaction density has always be considered as a rather elusive quantity, and a number of studies have discussed [63, 64] whether it could be amenable to experimental determination. The combined ΔLS profile studies sketched above should permit to individuate the causes and magnify the density changes leading to the interaction density. The preliminary investigation reported in [13] certainly deserves to be deepened.

3.4 SF Description of Metal–Metal and Metal–Ligand Bonds in Organometallics

3.4.1 Why Exploiting the SF for Organometallics?

The relationships between the geometrical and electronic structure of transition metal complexes and the description of how metal atoms get bonded to one another or to the ligands have been the source of lively and debated discussions in literature through the years and are still the subject of a growing number of ongoing theoretical and experimental investigations [2, 14, 60, 65–79]. The ligands customarily provide the necessary glue for the energy stabilization of transition metal complexes, and the study of the interplay and/or competition between metal–metal and metal–ligand bonding is an interesting and challenging subject.

For a number of reasons detailed below, the SF has enjoyed an increasing attention as a useful tool to be exploited in this area, and especially so when used in combination with other techniques. SF applications have in particular concerned

the metal–metal (M–M) bond and the metal–ligands (M–L) interactions in polynuclear transition or alkaline-earth metal compounds [14, 77, 80–82], and also the M–L interactions in mononuclear metal complexes involving π-bonded unsaturated hydrocarbyl ligands [72, 83] or σ-bonded ligands [84].

Topological approaches to the description of chemical bonding have revealed how many traditional bonding paradigms become no longer appropriate when applied to organometallics [85–87]. Moreover, it has become evident that the well-consolidated bonding classification schemes derived from the various topological approaches also need to be critically analyzed and even revised when metal–metal (M–M) bonds or metal–ligand (M–L) interactions are concerned [2, 66, 85, 86]. Problems involve in a first instance the choice of a suitable criterion to establish which atoms are actually bonded to one another – eventually leading to the so-called molecular structure – and, second, the characterization and classification of the resulting chemical bonds.

When translated to common "chemical thinking," the very successful and "universal" bond path criterion [7], which defines whether two atoms are bonded to one another, appears, in the case of organometallics, not completely free of limitations [2, 60, 66, 72]. For these systems, continuous rather than discontinuous bonding indicators and descriptions seem to be perhaps more appropriate. Very soft potential energy surfaces – hence often very flat electron densities – characterize the M–M and M–L interactions in organometallics, so that the resulting structural diagrams exhibit an enormous sensitivity to computational or experimental details [2, 66, 88]. And, as a consequence, the presence or lack of these interactions, when judged solely by the bond path criterion, may in some instance be a rather subtle and controversial issue. On top of this, structure diagrams of organometallics are often at or very close [66, 72, 80, 83, 88, 89] to bifurcation or conflict catastrophe points so that their structures typically travel from one structural region to another one just for very small displacements along one of their softer vibrational modes. As a challenging example, a system, whose structure diagram evolves for a number of reasons through a conflict mechanism by moving through points of the nuclear configuration space in the immediate neighborhood of the conflict catastrophe point, will exhibit two alternative pair of atoms competing for a bond path, despite the electron sharing for these two pairs of atoms –measured by the delocalization index – will be almost physically indistinguishable. However, according to the bond path criterion and the electron density topology, one pair will be termed as "bonded" and the other one as "not bonded." Even when the structure is not so topologically unstable, interpretation problems may arise. Gatti and Lasi [14] showed, for instance, that the electron sharing between the metal atoms in unsupported binuclear metal carbonyls may be comparable or even smaller than in the corresponding bridged carbonyl compounds, notwithstanding a bond path is found to connect the metal atoms in the former and not in the latter compounds. This observation is obviously tied to the intrinsic inability of the bond path criterion to directly detect multicenter bonding [60]. Bonding through the bridging ligands and direct M–M bonds are usually alternative and competitive options using the bond

path criterion [60], but they are not necessarily so when continuous descriptors of bonding, like the delocalization indices or other tools discussed below, are used.[13]

Within interacting quantum atoms (IQA) theory [90], for instance, this dichotomy may be interpreted [8] as a delicate case of exchange energy competition, where a bond path is found to link which of the two alternative atom pairs, M–M or M–L, has a dominant interatomic exchange energy; the existence of a possibly notable electron sharing and covalent interaction also between the "nonbonded" pair of atoms is not denied within this approach, but, rather precisely defined and quantitatively evaluated [8].

The domain-averaged Fermi hole (DAFH) analysis [91–93] is another very useful interpretive tool adopted in this area. When applied to the highly debated case of the triply bridged $Fe_2(CO)_9$ coordination complex, where the 18-electron rule would predict a direct Fe–Fe bond, the DAFH approach rather than this direct M–M interaction suggests the existence of a multicenter 3c–2e character of the bonding of the bridging ligands [60, 65]. This view nicely fits with the nonnegligible electron sharing found between the two metal atoms, despite the absence of a direct Fe–Fe bond, since the existence of nonvanishing delocalization indices between all pair of atoms has been proved to be a necessary requirement for the presence of 3c–2e bonding in any A–B–C fragment [94].

The ELF approach [54, 55], instead, distinguishes among different bonding schemes by assigning a synaptic order to each of the recovered ELF valence basins and by finding the number and type of core basins with which they have a boundary [95, 96]. For instance, disynaptic valence basins are associated with conventional two-center bonds and trisynaptic basins with 3c–2e bonds. Electron populations of such ELF valence basins then denote their hierarchical importance.

Another tool, able to overcome the problems inherent to the possibly discontinuous description of bonding provided by the electron density topology and the bond path criterion, is clearly the SF. Although lacking the very important physical meanings associated with either the delocalization indices or the IQA, DAFH, and ELF analyses, the SF has the great advantage of not requiring the pair density (or at least the first density matrix if single determinant theoretical approaches are used) for its application and for being so, as repeatedly mentioned in this review, immediately applicable to both experimental and theoretical electron densities. This is an important mark since until recently the most decisive features of the experimental studies of bonding in organometallics were often derived from complementary theoretical calculations – a situation that clearly raises the question of whether it is indeed worth performing the more time-consuming experimental determinations for such systems [14].

[13]It is worth noting that there is nothing wrong nor contradictory in the QTAIM description of these bonds; simply, different though complementary views emerge when, using such theory, information from the position space, where bond paths are made manifest, is combined with that derived from the six-dimensional pair density space, where electron sharing among atomic basins and electron localization within atomic basins take place and compete between themselves.

Once the important interactions in an organometallic structure have been detected, the problem arises of how to categorize them [2, 66]. The dichotomous classification [2, 97] of bonding interactions based on the sign of the Laplacian of the electron density at the bcp, $\nabla^2\rho_b$, is known to be inappropriate [2, 66] for bonding between atoms whose atomic $\nabla^2\rho$ distributions lack the outermost regions of charge depletion and concentration, as it is the case for most of the transition metal atoms. On top (and partly because) of this peculiar trait of the M atomic Laplacian distributions, M–M bonds are often characterized by a very low $\nabla^2\rho_b$ value, implying the practical indeterminacy of the $\nabla^2\rho_b$ sign and the adoption of the sign and value of $\nabla^2\rho_b$ as a sole classification index largely misleading [70, 71]. Other bond indices, like those related to the continuous descriptors mentioned earlier or those derived from the virial and kinetic energy densities at the bcp, e.g., the energy density $H_b = V_b + G_b$, [98], have often been used for organometallics [67, 71] and their performance on this kind of compounds critically analyzed [2, 14, 66]. However, these indices are also not directly amenable to experimental determination. They are customarily derived [99] through an approximate formula from DFT theory using the experimental density, its gradient and Laplacian and can thus hardly be regarded as true experimental outcomes. In addition, questionable appears the reliability of the approximated values of such indices when bonding interactions characterized by very different values and consequently a very distinct nature are scrutinized. Indeed, if the "exact" quantum mechanical values for V_b and G_b are compared with the corresponding estimates from the approximate DFT formula, one typically observes a very unbalanced replica of the exact data, with values for closed-shell interactions being in an almost quantitative agreement and those of typical shared interactions being very poorly, if not at all reproduced [2, 44, 99].

Another increasingly used path to the *experimental* determination of quantities derivable from a single-determinant wavefunction (including, *inter alias*, delocalization indices, V_b, G_b, exchange energies, ELF and DAFH distributions, etc.) is that of obtaining a wavefunction fitted to experiment by constraining the theoretical wavefunction of a model system for the crystal to reproduce the crystal experimental structure factor data within a given accuracy [100, 101]. Despite the relevance and merits of such an approach, one cannot deny that substantial theoretical contamination may affect the experimental model wavefunction and the quantities derived thereof. Not less importantly, the practical application of the constrained wavefunction method implies some kind of arbitrariness, in particular in the choice of a suitable model system for computing the theoretical wavefunction and in the selection of the optimal weight to be given to the experimental data when constraining the wavefunction to such data.

All in all, exploiting the SF for organometallics meets three very important prerequisites: (a) to take advantage of a tool capable of detecting both proximal and distant contributions to a chemical bond, thus overcoming the restrictive two-center view of bonding; (b) to adopt a continuous rather than a discontinuous descriptor of bonding interactions, hence apt to analyze them also in systems often characterized by topological structural instabilities; (c) to make use of an approach which is

applicable on the same ground to both experimental and theoretical electron densities and without any approximation on either side.

3.4.2 Metal–Metal and Metal–Ligands Interactions in Polynuclear Transition Metal and in Alkaline-Earth Metal Compounds

Gatti and Lasi [14] have applied the SF tool to a set of saturated [$M_2(CO)_x$, M= Mn, Fe, Co, Ni, x = 10–7] and unsaturated [$Co_2(CO)_x$, x = 8–5] binuclear 3*d* metal carbonyls, and to the M_2(formamidinate)$_4$ (M=Nb, Mo, Tc, Ru, Rh, and Pd) binuclear 4*d* metal complexes, using ab initio electron densities. Their study focused on a number of questions listed below:

(a) How the SF describes the rather elusive M–M bonding interaction in the 3*d* series and how differentiates this interaction from the supposedly stronger M–M bonding in the 4*d* binuclear metal complexes?

(b) Does the SF reveal any difference between the M–M bonding in the saturated and in the variedly unsaturated 3*d* binuclear metal carbonyls? If it does, how this relates with changes in the formal bond order?

(c) How does the SF description compare with that provided by the usual local bond indices or by more sophisticated descriptors such as the delocalization indices or the ELF?

(d) Can the SF distinguish the case of a direct M–M interaction from that of an indirect metal–metal bond occurring through the bridging ligands?

Selected answers to such questions are summarized in the following, through a number of worked examples.

Figure 8 (top) shows the atomic SF percentage contributions to the Mn–Mn, $(Mn–C)_{eq}$, and $(CO)_{eq}$ bcps in $Mn_2(CO)_{10}$.[14] Comparison of sources displayed in Fig. 8 immediately reveals the profound difference in nature among these bonds.

While the dative Mn–C bonds and in particular the stronger C–O bonds have fairly localized or very localized sources, with percentage contributions from the two bonded atoms to the bcp density equal to 67% and 97–98%, respectively, sources for the Mn–Mn interaction look dramatically different. Contrary to all bonds examined thus far, the two Mn bonded atoms yield a negative contribution (−50%) to the bcp density, and the required positive ρ_b value is determined by the largely outweighing positive contributions from the carbonyl groups linked to the metal atoms (12.5% and 15.7% from each axial and each equatorial group, respectively). Even in the case of the non-covalent OH···O bond, the negative source from the H involved in this bond is already compensated for by the positive source from the H-acceptor atom or, for the weakest bonds, when this source is summed up to that from the H-donor atom. Hence, while the three atoms directly involved in the

[14]Data for the axial Mn–C and C–O bonds are reported in [14] and are negligibly different from those for the corresponding equatorial bonds, with changes which closely parallel the observed bond length differences.

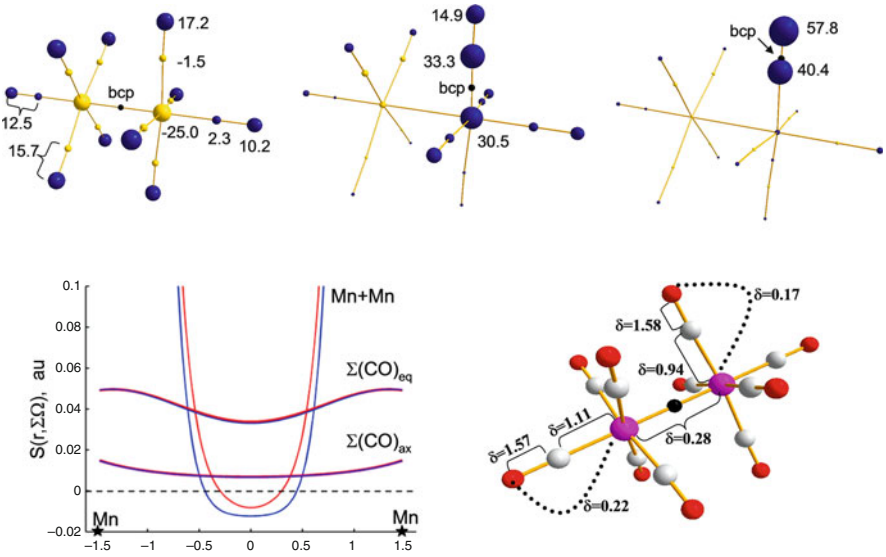


Fig. 8 Bis(pentacarbonylmanganese) $Mn_2(CO)_{10}$. *Top*: Source function percentage contributions to the M–Mn (*left*), Mn–C_{eq} (*middle*), and (C–O)$_{eq}$ (*right*) bcp densities. *Bottom*: Source function profiles (*left*) along the Mn–Mn internuclear axis for $Mn_2(CO)_{10}$ (*blue lines*) and the model system composed by two noninteracting $Mn(CO)_5$ fragments (*red lines*); delocalization indices (*right*). In part and adapted from Figs. 4 and 5, with permission from [14], http://pubs.rsc.org/en/Content/ArticleLanding/2007/FD/b605404h Copyright 2007, The Royal Society of Chemistry (RSC)

H-bond always suffice to ensure a positive contribution to the HB bcp density, it is the sum of contributions from all atoms other than the bonded ones that ensures the necessary positive density at the Mn–Mn bcp in $Mn_2(CO)_{10}$. This result neatly complies with the dominant role the ligands have in causing the two metal atoms to be bonded to one another and characterizes the M–M interaction as a multicenter one or, perhaps more correctly, as a one *determined* by a multicenter synergic cooperation.

Interaction with the carbonyl ligands polarizes the metal atomic Laplacian distribution in such a way that the metal determines a subtraction rather than a positive contribution to the density at the Mn–Mn bcp. Inspection of the profiles of the SF (Fig. 8, bottom left) and of the LS (Fig. 9) along the Mn–Mn bond path enables one to have a deeper understanding on how this negative density contribution at the Mn–Mn bcp from the Mn atoms originates.

In Fig. 8 (bottom left), profiles of the SF contributions from the metal atoms and from the two axial and the eight equatorial carbonyl groups are compared for $Mn_2(CO)_{10}$ (blue curves) and a model system made by two noninteracting Mn $(CO)_5$ fragments (red curves), placed in the same geometry as in $Mn_2(CO)_{10}$. Noninteracting fragments mean that the electron density rearrangement due to the Mn–Mn bond formation has been *switched off*, while each fragment retains its DFT self-consistently determined electron distribution. The SF contribution from the

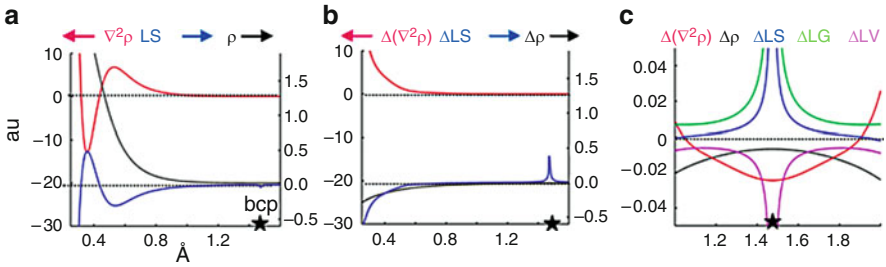


Fig. 9 $Mn_2(CO)_{10}$. (**a**) Profiles of the local source (LS), $\nabla^2\rho$ and ρ along the bond path; (**b**) difference Δ of the profiles in $Mn_2(CO)_{10}$ and in the model system composed by two noninteracting $Mn(CO)_5$ fragments for the same properties shown in the *left panel*; (**c**): difference profiles as in (**b**), but magnified by a factor of 100 around the bcp. ΔLG and ΔLV are differences in the kinetic and potential energy local source contributions profiles. For LS, LG, and LV properties, the bcp is selected as reference point [partly adapted from Fig. 6, with permission from [14], http://pubs.rsc.org/en/Content/ArticleLanding/2007/FD/b605404h Copyright 2007, The Royal Society of Chemistry (RSC)]

carbonyl groups remains always positive and smoothly varying along the whole bond path with the contributions from the axial and equatorial carbonyls roughly paralleling their different numerical consistency and with no significant changes between the real and the model system (the blue and the red lines are almost superposed). On the other hand, the SF contributions from the metal atoms – that largely overcome those from the carbonyls up to 0.9 Å from the nuclei – are shown to decrease steeply on moving to the Mn–Mn midpoint and to become even negative in a large interval of about 0.8 Å around such point. The effect is more evident when the Mn–Mn bond formation is *switched on*, even though the polarization due to the ligands in the $Mn(CO)_5$ fragments already suffices to yield negative, yet smaller SF contributions from the Mn atoms in a narrower interval around the midpoint.

Analysis of the local source (LS) profiles introduces further detail and provides additional insight. Figure 9 displays (panel *a*) LS, ρ, and $\nabla^2\rho$ profiles along the Mn–Mn bond path for $Mn_2(CO)_{10}$ and (panel *b*) their differences ΔX (X = LS, ρ, and $\nabla^2\rho$) with respect to the model system with no Mn–Mn bond. The atomic Laplacian distribution for the fourth row atoms from Sc to Ge – thus comprising Mn – lacks the outermost *N*-shell of charge depletion and concentration, and the bcp in $Mn_2(CO)_{10}$ falls in the region of charge depletion (CD) of the *M* atomic shell [14]. Hence, the LS (with the Mn–Mn bcp taken as rp) becomes already negative at about 0.46 Å from the Mn nucleus, where the region of charge concentration of the *M* shell has its end, and so provides negative local contributions to the density at the bcp for a very large interval of the bond path. The same holds true for a model system composed by two noninteracting Mn atoms placed at the same distance as in $Mn_2(CO)_{10}$ (Fig. 6a in [14]). However, we know that each of these noninteracting Mn atoms has to necessarily yield a positive SF contribution to the density at the Mn–Mn midpoint (mp), equal to 0.5 ρ_{mp}, since no other atoms are present in the system. When the LS profile for this model system is compared to that for

$Mn_2(CO)_{10}$, it turns out that the negative $S(\mathbf{r}_{mp},Mn)$ value in $Mn_2(CO)_{10}$ is clearly the result of a notably shrunk *M*-shell charge concentration region and of significantly less negative Laplacian values within this region. Similar changes, though slightly less evident, are found when the comparison is made with respect to the model system of the two noninteracting $Mn(CO)_5$ fragments (Fig. 9b). All in all, when the Mn–Mn bond formation is enabled, the *M*-shell charge concentration (CC) region becomes smaller and less concentrated, so becoming less effective in producing a final positive source contribution at the Mn–Mn mp with respect to the two studied model systems. This result seems to contradict the expected charge concentration increase, in the metal–metal internuclear region, when Mn–Mn bond is enabled. Inspection of Fig. 9c, where difference profiles around the bcp are magnified by a factor of 100 with respect to those in Fig. 9b, *solves this apparent dilemma*. Even though enabling the Mn–Mn interaction affords a negative chemical deformation density along the whole bond path ($\Delta\rho$ remains always negative), $\Delta\nabla^2\rho$ – that was largely positive in the *M* shell concentration region – becomes now slightly negative. In a region of about 1 Å around the bcp, charge is thus less depleted in the bonded system than it is in the model system where Mn–Mn bonding is disabled. At midpoint, the Laplacian of the density is one order of magnitude less positive in $Mn_2(CO)_{10}$ ($\nabla^2\rho_b = 0.3 \times 10^{-2}$ a.u.) than it is in the model system made by the two $Mn(CO)_5$ fragments [$\nabla^2\rho(mp) = 2.7 \times 10^{-2}$ a.u.]. Accordingly, ΔLS becomes positive in this region, indicating that when the Mn–Mn bond is enabled, less charge is subtracted from the mp density by this region.[15] This result may be further dissected in terms of the local kinetic and potential energy density contributions to the local sources for the two analyzed systems. The less negative LS of $Mn_2(CO)_{10}$ around the bcp is the result of a larger decrease in the kinetic energy density as compared to the parallel decrease in the magnitude of the potential energy density in this system.[16] As a consequence, ΔLG is more positive than ΔLV is negative around the bcp,[17] so disclosing the physics behind the less negative local source contribution when the metal–metal bonding is allowed. The less positive Laplacian, less negative LS, and the higher local dominance of the potential energy density [|V|/G (mp) = 1.92 in $Mn_2(CO)_{10}$ and 1.53 in the model system] are all signs of an increased *local* electron pairing or covalency when bonding is permitted. In conclusion, the following two opposing factors come to play when bonding is enabled, as for the SF contribution from the Mn atom to the Mn–Mn bcp density: (a) the *M*-shell charge concentration region shrinks in volume and decreases its average charge concentration and (b) the electronic charge becomes less depleted in a vast region around the bcp. The more negative S(Mn)

[15]The ΔLS spike at the Mn–Mn mp results from $\nabla^2\rho$ being one order of magnitude greater in $Mn_2(CO)_{10}$ at this point than in the model system with no Mn–Mn bond.

[16]G and |V| are lower in $Mn_2(CO)_{10}$ than in the model system with no metal–metal bond because ρ at mp is lower in the former than is in the latter system (2.6×10^{-2} vs. 3.3×10^{-2} au).

[17]Remember that LG or LV are both related to the negative of the kinetic or potential energy densities; hence, ΔLG and ΔLV have opposite sign with respect to the differences of the kinetic or potential energy densities in the $Mn_2(CO)_{10}$ and in the system with no Mn–Mn bond.

values for $Mn_2(CO)_{10}$ with respect to those of the model system where the bond between the two Mn atom is hindered (Fig. 8, bottom left) simply reveal that the decrease in the charge depletion around the bcp is not large enough to compensate for the effects of the accompanying charge concentration decrease in the *M* shell of the metal. Analysis of the LS profiles enables one to dissect the diverse modifications occurring in the various atomic shell regions of the metal when the metal atoms get bonded to one another and to disclose how these local changes affect the overall metal source contribution to the density at the bcp.

Figure 8d displays delocalization index values δ showing how electrons are shared among vicinal and 1,3 atomic pairs in $Mn_2(CO)_{10}$. These indices were already introduced in Sect. 3.2.1, where the description provided by the SF and by the δ in a number of prototypical systems was compared, and where the necessary caveats related to this kind of comparison were pointed out. Keeping in mind that delocalization indices and source function contributions are not physically related in a direct way, it is yet worth noting a number of evident correspondences. The small magnitude of the S%(bcp,Mn) value complies with the very low δ(Mn,Mn′) value of 0.28 (Table 9), which is clearly quite far from that expected for a formal single bond.[18] It is rather comparable to that found between Mn and either the axial or the equatorial oxygen atoms ($\delta = 0.22$ and 0.17 respectively), which are both only 1,3 indirectly bonded to Mn through their corresponding C atoms. The relatively large number of electrons shared between the Mn atom and the carbonyl O atoms goes with the important S% contributions from the O atoms to the Mn–Mn′ bcp density. As discussed earlier, Mn–C and C–O bonds are characterized by much higher S% contributions from the bonded atoms than is for the bond between metal atoms and so, not surprisingly, the $\delta(\Omega, \Omega')$ values are close to 1 and to about 1.6 for the Mn–C and the C–O bonds, respectively. Moreover, the δ for the equatorial or axial bonds is found to be ordered in value as are the sums of the percentage SF contributions from the two bonded atoms [14].

Table 9 summarizes M–M′ bond properties for the binuclear 3*d* and 4*d* metal complexes mentioned at the beginning of this section. Metal complexes in this table and from now on in this section are identified as M.*x*.FBO.nb, where M is the metal, *x* the total number of ligand in the complex, FBO the formal bond order based on the 18-electron rule, and nb the number of bridged ligands (nb = 0 for an unbridged system). For instance, $Mn_2(CO)_{10}$ and $Fe_2(CO)_9$ are thus denoted as Mn.10.1.0 and Fe.9.1.3, since both have a formal bond order of one but 0 and 3 bridging ligands, respectively.

Among the saturated binuclear 3*d* metal carbonyls, only the two unbridged compounds Mn.10.1.0 and Co.8.1.0 exhibit an M–M bcp, despite all systems share the same formal bond order of one. Features of metal–metal bonding in the two unbridged compounds are qualitatively alike, with the M atoms in both

[18]The fact that S(bcp,Mn) is not only small but also negative is due to the polarization of the $\nabla^2\rho$ distribution into the Mn basin induced by Mn–Mn bonding and to a minor extent to bonding to ligands, as shown earlier in this chapter. No special relationship may be sought instead between a negative S% value and δ, which is necessarily ≥ 0 by definition.

Table 9 Topological metal–metal (M–M′) bond properties, source function percentage contribution from metals to the M–M′ bond critical point, and M–M′ delocalization indices in a number of binuclear 3*d* and 4*d* metal complexes[a]

M.*x*.FBO.nb[b]	$R_{M-M'}$ Å	$\rho_b \cdot 10^2$	$\nabla^2\rho_b \cdot 10^2$	$\lvert V_b\rvert/G_b$	G_b/ρ_b	$H_b \cdot 10^2$	H_b/ρ_b	q(M)	S% (M)	δ(M,M′)
Saturated binuclear 3*d* metal carbonyls, $M_2(CO)_x$										
Mn.10.1.0	2.949	2.62	0.27	1.91	0.29	−0.71	−0.27	0.952	−25.0	0.285
Fe.9.1.3	2.515	4.80*[c]	9.30	1.36	0.77	−1.34	−0.28	0.862	−5.2	0.355
Co.8.1.2	2.548	4.61*[c]	8.72	1.37	0.75	−1.29	−0.28	0.705	0.6	0.356
Co.8.1.0	2.691	3.61	0.20	1.95	0.35	−1.19	−0.33	0.658	−4.4	0.437
Ni.7.1.1	2.663	3.70*[c]	6.21	1.36	0.65	−0.86	−0.23	0.609	4.5	0.226
Unsaturated binuclear Co metal carbonyls, $Co_2(CO)_x$										
Co.7.2.0	2.486	4.94	3.77	1.67	0.57	−1.89	−0.38	0.636	7.3	0.394
Co.7.2.1	2.394	5.79*[c]	8.55	1.50	0.74	−2.17	−0.37	0.645	14.6	0.783
Co.6.3.2	2.254	7.19	15.31	1.40	0.90	−2.61	−0.36	0.616	18.5	0.726
Co.5.4.1	2.168	7.85	23.56	1.31	1.09	−2.70	−0.34	0.530	27.9	0.976
M_2(formamidinate)$_4$ binuclear 4*d* metal complexes, $M_2(HNCHNH)_4$										
Nb.4.3.4	2.284	13.00	35.46	1.37	1.07	−5.09	−0.39	1.413	40.4	2.507
Mo.4.4.4	2.141	16.57	51.79	1.36	1.22	−7.33	−0.44	1.360	40.6	2.930
Tc.4.3.4	2.122	16.40	50.16	1.36	1.20	−7.08	−0.43	1.208	40.8	2.713
Ru.4.2.4	2.540	7.32	6.59	1.63	0.61	−2.81	−0.38	1.042	28.2	1.316
Rh.4.1.4	2.501	7.27	6.93	1.61	0.61	−2.71	−0.37	0.962	29.3	1.035
Pd.4.0.4	2.730	4.16	14.99	1.12	1.04	−0.53	−0.13	0.815	17.1	0.294

[a]Data from [14]. If not otherwise stated, all quantities in au. V_b, G_b, and H_b are the potential energy density, the positive-definite kinetic energy density, and the energy density at the M–M bcp or at the M–M midpoint (see point c)

[b]Each studied system is denoted as M.x.FBO.nb, where M is the metal, *x* the total number of ligands in the complex, FBO the formal M–M bond order based on the 18-electron rule, and nb the number of bridged ligands, with nb = 0 in the case of an unbridged system. For the isostructural M_2(formamidinate)$_4$ binuclear 4*d* metal complexes, *x* and nb are fixed to 4, while the FBO varies from 0 in the Pd compound to 4 in the Mo complex. For those systems, Ni.7.1.0 and Co.7.2.0, with two nonequivalent M atoms, averaged M atom values are reported for FBO, *q*(M) and S%(M)

[c]The asterisk, if present, denotes metal complexes lacking an M–M bcp. If this occurs, properties are evaluated at the M–M midpoint or, for the S%(M) value, using this point as the rp

compounds determining a negative contribution to the density at bcp and with the carbonyl groups largely compensating such a density subtraction from the bcp. The smaller negative S%(bcp,M) in Co.8.1.0 complies with a larger δ(M, M′) value (0.437 vs. 0.285) and a smaller net charge on the metal (q(M) = 0.658 vs. 0.952) in this compound. The net charge on the M atom decreases along the whole series with the decrease in the number of linked ligands, and the S%(bcp or mp,M) values correspondingly increase, when the trends for the bonded and not bonded systems are examined separately. In fact both the ligands and, in particular, the M–M bonding concur in decreasing the S%(M)[19] value, as discussed earlier by comparing

[19]S%(M) is here used for either S%(bcp,M) or S%(mp,M) according the system has or not has an M–M bcp.

the Mn.10.1.0 system with the two model systems where either M–M bonding or also the effect of the carbonyl ligands was disabled. The metal sources for the M–M bonded systems are comparatively more negative because of the enhanced decrease in the charge concentration within the metal atomic M shell due to M–M bonding.

It is worth emphasizing, however that, irrespective of the lack or presence of a bcp, the picture of a nonlocalized M–M interaction with features quite different from those of a conventional covalent bond emerges. The electron sharing between metals is small through the whole series [δ(M, M′) =0.29–0.44] and quite far from one, and the S%(M) values are also small, with the electron density at the M–M bcp or mp being largely determined by atoms other than the metals. As anticipated earlier, the nonvanishing δ(M,M′) values for the systems with no M–M bond is a necessary consequence of the indirect 3c–2e M–M bonding through the bridging ligands [60, 94].

The unbridged (D_{3d}, Co.8.1.0) and the double-bridged (C_{2v}, Co.8.1.2) $Co_2(CO)_8$ isomers offer the possibility to compare direct and indirect M–M bonding features for two systems having the same metal atoms and number of CO ligands. Figure 10 shows that for both systems the electron density at bcp or at the Co–Co mp is essentially determined by the carbonyl groups and in particular by the carbonyl

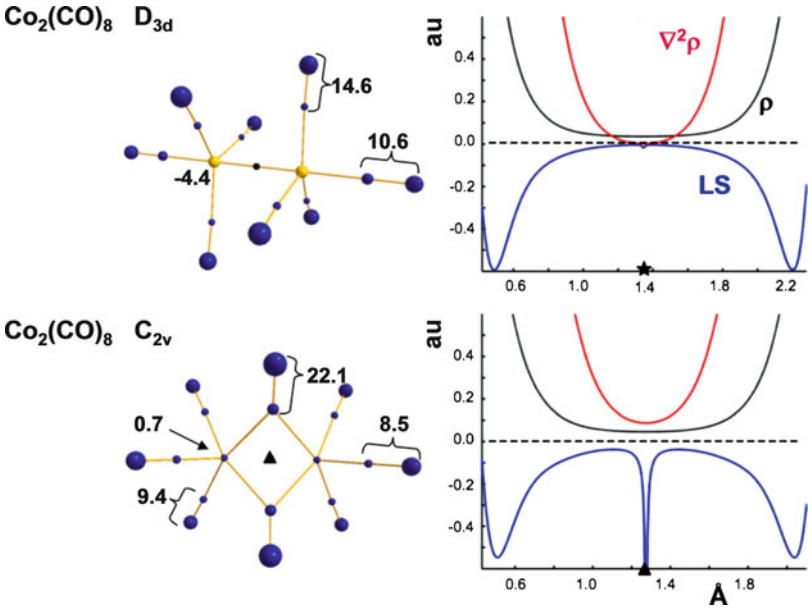


Fig. 10 Source function percentage contributions at the Co–Co midpoint (*left*) and local source (LS), $\nabla^2\rho$ and ρ profiles along the Co–Co internuclear axis (*right*) for the unbridged D_{3d} (*top*) and the double-bridged C_{2v} (*bottom*) $Co_2(CO)_8$ isomers. Bcp and midpoint locations are denoted by a star and a triangle, respectively. The Co–Co bcp or midpoint is used as reference point for the LS profiles [adapted from Fig. 7, with permission from [14], http://pubs.rsc.org/en/Content/ArticleLanding/2007/FD/b605404h Copyright 2007, The Royal Society of Chemistry (RSC)]

oxygen atoms. However, the relative contributions from these groups clearly vary between the two isomers, as the two bridging carbonyl groups in the C_{2v} form bring a notably enhanced S% contribution, equal to 44.2%, and almost comparable to that from all other six carbonyl ligands (S% = 54.6). In the unbridged isomer, instead, the average contribution from each carbonyl group amounts to 11.6% and with no large difference between contributions from each axial (S% = 14.6) or equatorial carbonyl group (S% = 10.6). Both the direct and the indirect M–M bonds exhibit a delocalization of sources for the density at the Co–Co mp, but the largely enhanced source from the bridging ligands in Co.8.1.2 clearly illustrates how the SF tool reflects the special role these specific ligands have into the indirect M–M bonding.

Despite the lack of a Co–Co bcp, the electron density at the Co–Co mp is higher in the bridged (4.6×10^{-2} au) than is in the Co–Co bonded isomer (3.6×10^{-2} au), as likely the result of the shorter Co–Co distance in the former (Table 9). However, direct metal bonding leads to a small increase in the electron sharing in the unbridged isomer [compare δ(Co, Co′) values in Table 9], despite the 0.15 Å increase in the Co–Co separation.

As discussed earlier at length for $Mn_2(CO)_{10}$, polarization of the metal atom distribution induced by M–M bonding leads to a S(Co) value slightly larger in the bridged ($+0.03 \times 10^{-2}$ au) than is in the Co–Co bonded isomer (-0.7×10^{-2} au), where it becomes even a little negative. Yet, analysis of the profiles of $\nabla^2\rho$ and of the LS along the Co–Co internuclear axis (Fig. 10) reveals complementary and quite distinct features for the direct and indirect M–M bonding cases. For both isomers, the Laplacian remains positive in a very large interval (>2 Å) around the Co–Co mp, and hence the LS referred to mp turns out to be there everywhere negative for either systems. However, since in a smaller interval of about 0.8 Å around the midpoint, $\nabla^2\rho$ is about one order of magnitude smaller in the unbridged than is in the bridged system, and the two isomers largely differ as for their corresponding negative contributions to the mp electron density. The LS profiles for the two isomers are thus clearly distinct around the Co–Co mp (Fig. 10). Both profiles are traced within the same distance from the mp, which is obviously a singular point for the LS, so that their different forms just magnify the difference in the Laplacian profiles for the two isomers around the midpoint. The more positive Laplacian for the isomer with no Co–Co bond path is mirrored in a LS sudden drop around the Co–Co mp rather than in the smooth LS curve observed for the bonded isomer.[20] This difference may be interpreted as a sign of larger local covalency in the unbridged isomer, despite its larger Co–Co separation and the somewhat lower electron density values along the internuclear axis (Fig. 10). The very small and not dissimilar percentage source contribution from the Co atom to the ρ value at the Co–Co mp in the two isomers results from a compensation. Regions close to

[20]Note that these LS profile comparisons are meaningful *only on a relative basis and when the* LS *are plotted within the same distance from the rp*! Decreasing this distance could make the bonded system also exhibit a clear drop in LS, since this function has a singularity at the rp. This warning is given here because some confusion has recently arisen in the literature as for such a use of the LS to discuss bonding in delicate cases.

the Co–Co mp determine more negative LS contributions to the mp density in the bridged than in the unbridged isomer, whereas the opposite holds true at the extremes of the LS profiles shown in Fig. 10. Indeed, far from the mp, the region of charge depletion related to the atomic *M*-shell is a bit more locally depleted in the unbridged system, so determining slightly more negative sources contributions from this region and for this isomer at the Co–Co bcp.

All in all, combined information from the integrated and local form of the source function discloses notable and interesting differences in the direct and indirect schemes of metal–metal bonding, differences which may be hidden or even lead to apparently contradictory results when descriptors like the electron density at the M–M mp, the S(M) or S%(M) values, and the δ(M,M′) values are examined separately.

The $[Co_2(CO)_x, x = 7–5]$ series covers the case of unsaturated binuclear 3*d* metal carbonyls, having the same metal atom and where the number of CO ligands decreases with increasing formal bond order from two to four through the series (Table 9). By including also the two saturated $Co_2(CO)_8$ isomers, we can review a set of systems having a formal bond order extending from one to four. All systems exhibit a Co–Co bcp, except the bridged ones with $x = 7$ and 8 (Co.7.2.1 and Co.8.1.2). The presence of bridging ligands does not instead inhibit the formation of a Co–Co bcp for the two systems with highest bond order and shortest bond length, Co.6.3.2 and Co.5.4.1.[21]

The S%(Co) for the density at the Co–Co bcp increases from −4.4% in Co.8.1.0 to a maximum of 27.9% in Co.5.4.1, on passing from a formal bond order value of one up to four. The S%(Co) value regularly increases through the series with the decrease in the net positive charge on the Co atom and of the number of CO ligands (Table 9). The SF contribution from the two Co atoms exceeds that from the ligands only for the highest formal bond order (55.8% and 44.2%), whereas for a formal bond order of two or even three, the contribution from the ligands largely dominates [89% on average for the two considered $Co_2(CO)_7$ systems and 63% for Co.6.3.2]. The S% from the two bonded atoms in Co.5.4.1 *is still far even from that of a standard single covalent bond, despite a formal bond order of four*. The peculiar nature of the M–M interaction in the 3*d* binuclear metal carbonyls anticipated by the SF analysis is corroborated, also for the unsaturated series, by the Co–Co delocalization indices. The number of shared pairs of electrons does not reach one even for the system with a formal bond order of four where it amounts to 0.976 pairs only, instead of four. Fine details as for the almost paralleling trends of S%(Co) and δ(Co, Co′) values through the series are discussed in [14]. Here, we emphasize the interesting result that a description of Co–Co bonding, thoroughly at variance with that provided by the 18-electron rule, can be already obtained using a tool which is based only on the electron density and on the electron density Laplacian.

[21]Note that also in the case of the bridged $Co_2(CO)_8$ isomer a Co–Co bcp occurs when the Co–Co distance is decreased by about 0.2 Å with respect to equilibrium [60].

We examine now briefly how the local bond properties customarily used within QTAIM describe the saturated and unsaturated binuclear 3*d* metal carbonyls listed in Table 9, and how this description confronts with or complements that provided by the SF. We consider first the saturated systems. As anticipated earlier, all show a positive $\nabla^2\rho$ value at the M–M midpoint, but with magnitudes one to two order of magnitude lower for the metal–metal bonded systems. Also the kinetic energy density per electron, G/ρ, and the $|V|/G$ ratio at the M–M bcp or mp clearly distinguish the M–M bonded from the M–M indirectly bonded systems, with G/ρ, having significantly lower and $|V|/G$ notably higher values for the former compounds. Note that the $|V|/G$ ratio predicts all systems in the series to fall in the *transit region* for incipient covalent bond formation, according to the proposal made by Espinosa et al. [102] on the basis of a systematic topological study of a set of neutral, positively and negatively charged X–H···F–Y gas-phase systems, and combined with a natural bond order (NBO) analysis of their wavefunctions. Assuming that an NBO analysis on M–M bonds would confirm the same interval, $1 \leq |V|/G \leq 2$, for the transit region of incipient covalent bond formation,[22] one observes that the $|V|/G$ value in Mn.10.1.0 and Co.8.1.0 is close (1.91 and 1.95) to the border with the shared-shell region, whereas it is not so far from the border with the closed-shell region (1.36–1.37) for the compounds without a M–M bcp. Analogously to the percentage source function contribution from the metal S(M), and to the delocalization index $\delta(M,M')$, the energy density, H, and in particular the energy density per electron, H/ρ, appear as unable to neatly distinguish direct from indirect M–M bond. Both quantities are negative, denoting a (partial) covalent character for the M–M interaction, regardless of its direct or indirect nature. H/ρ, also called *covalence degree* [102], remains almost stable through the series, whereas H slightly oscillates around -1×10^{-2} au and reaches the highest values for two out of the three bridged systems, possibly because of their shorter M–M distance. A similar behavior is observed for the ρ values at the M–M bcp or midpoint location. In summary, in the case of saturated systems, a number of bond properties at this location, such as $\nabla^2\rho$, G/ρ, the $|V|/G$ ratio, and the LS (rp ≡ mp) seem to be associated with the local extent of electron sharing between the metal atoms, whereas the S%(M), the (M,M') delocalization indices, the H, and H/ρ descriptors appear to be more related to a cumulative estimate of such sharing and regardless of whether it takes place directly or through the bridging ligands. Combined use of the SF and of the LS tools enables one to merge these two complementary descriptions in a unified picture, as also already shown for the two $Co_2(CO)_8$ isomers.

[22]To be rigorous, the transit region as defined by Espinosa et al [102] has thus far been proved to correspond to the bonding molecular orbital formation only for the H···F interaction. For other pairs of interacting atoms, the interval where this association takes place could be more or less displaced toward either the shared shell or the closed-shell regions, according to the specific nature of the interacting atomic pair. One anticipates that the above correspondence should hold true for chemically related interactions such as the NH···O, OH···O and CH···O hydrogen bonds, but, for instance, not necessarily for M–M bonds.

In the case of unsaturated systems, the local property values listed in Table 9 exhibit quite *unexpected* trends versus the formal bond order. To include the case of a formal bond order of one, the Co.8.1.0 saturated system with a Co–Co bcp and formal bond order of one is added to the unsaturated series [$Co_2(CO)_x$, $x = 7$–5] in this part of the discussion. The $\nabla^2\rho$ value is always positive and regularly increases with increasing formal bond order. Despite being at odds with the dichotomous classification based on $\nabla^2\rho$, this trend is clearly anticipated. For all systems, the bcp or mp falls within the *M*-shell charge depletion region of the isolated atom, and the Laplacian past the maximum of depletion in this region has to decay monotonically down to zero with increasing distance from the nucleus. The shorter the Co–Co distance is, the more positive will thus be the Laplacian at the Co–Co mp or bcp. The bond classification index based on the |V| /G ratio decreases monotonically along the unsaturated series reaching a value of 1.31 in the system with formal bond order of four. The largest value of 1.95 is instead observed for the saturated bonded Co.8.1.0 isomer with formal bond order of one. This is clearly the opposite trend one would have expected from this descriptor, which appears to depict as very close to a shared interaction the system with a formal bond order of one and as close to the closed-shell regime that with a formal bond order of four. On top of this, all systems in the [$Co_2(CO)_x$, $x = 8$–5] series are described as lying in the transit region for incipient covalent bond formation, despite the noticeable decrease in the Co–Co bond distance through the series. The bond degree, H_b/ρ_b, is almost constant through this series, with almost equal values for the first (−0.33) and the last (−0.34) term of the series. The kinetic energy per electron, G_b/ρ_b, monotonically rises along the series, as found also for standard covalent bonds (0.23, 0.39, and 0.68 in ethane, ethene, and ethyne). However, the last term of the series has a G_b/ρ_b value greater than one, which is normally regarded [6] as a sign of a closed-shell interaction or of an interaction with large polarization and/or charge transfer (which is obviously not the case here). Taken as a whole, use of standard local descriptors for the M–M bonding in the unsaturated [$Co_2(CO)_x$, $x = 8$–5] series appears problematic and leading to apparent contradictions with respect to the known behavior of such indices for bonds between atoms of the first three rows of the periodic table. Conversely, the S%(M) and the δ(M,M′) values lead to a qualitatively similar and meaningful description, which features a notable departure from the formal bond order classification based on the 18-electron rule model.

The M_2(formamidinate)$_4$ (M = Nb, Mo, Tc, Ru, Rh, and Pd) binuclear 4*d* metal complexes form a series of isostructural compounds which differ only for the nature of their metal atom and exhibit formal bond orders varying from zero (M=Pd) up to four (M=Mo) as a function of the number of electrons available for the M–M bond. The notation we adopt for these tetra-bridged systems, which have a paddlewheel structure, is analogous to that used for the 3*d* binuclear metal carbonyls. For instance, Ru.4.2.4 will denote the M_2(formamidinate)$_4$ complex with M = Ru and formal bond order of 2. The number of ligands, four, and their bridging nature are clearly fixed through the series.

The study of these complexes enables us to examine M–M bonding between 4*d* metal atoms and to observe whether this bonding has distinct features from that

between M atoms in the binuclear 3*d* metal carbonyls. Full details of this study may be found in [14], while only the main results are reviewed here. As shown in Table 9, each complex exhibits an M–M bcp, even the one with M = Pd and a zero formal M–M bond order. Furthermore, the ρ_b, S(M) and S%(M) values are much larger than those found for the M–M bond in the saturated binuclear 3*d* metal carbonyls or in the unsaturated binuclear Co carbonyls series with corresponding M–M formal bond orders. The Mo atom determines more than 40% of the density at the Mo–Mo bcp in Mo.4.4.4, analogously to what observed for the bonded atoms in a standard covalent single bond between second-row atoms, and definitely more than the contribution (27.9%) the Co atom yields to the Co–Co bcp density in $Co_2(CO)_5$, despite both systems do share an equal M–M formal bond order of four. The sum of the S% contributions from the two M atoms in Rh.4.1.4 almost reaches 60, while that for the binuclear metal carbonyls having the same formal bond order of one oscillates around zero and is even largely negative in Mn.10.1.0.

The notably larger S%(M) contribution to the M–M bcp density in the M_2(formamidinate)$_4$ (M = 4*d* transition metal) compounds relative to that in the saturated/unsaturated binuclear Co carbonyl series with corresponding M–M formal bond orders fits with the documented enhancement in the M–M bond strength down a group, arising from the greater spatial extension of *d* orbitals in heavier atoms [103]. For the bulk metals, the M–M bonds in the *d*-block are strongest in the 4*d* and 5*d* series, and this feature is known to carry over into their compounds [103]. In nice compliance with the largely enhanced S%(M) values, also the metal–metal delocalization indices for the M_2(formamidinate)$_4$ series are much greater than for the cobalt carbonyls series (Table 9). For Mo.4.4.4, δ(Mo,Mo′) amounts to 2.930, a value about three times as large as that found [δ(Co,Co′) = 0.976] for the binuclear Co compound sharing the same formal bond order of four, while Rh.4.1.4, having a formal bond order of 1, shows a δ(Rh,Rh′) value very close to one to be compared with an average value of about 0.4 for the two $Co_2(CO)_8$ isomers with analogous formal bond order. It is worth noting that the δ(M,M′) values are ordered as the formal bond orders for the M_2(formamidinate)$_4$ series, whereas the corresponding S%(M) values do not seem to clearly distinguish the compound with formal bond order of four (M = Mo) from the two systems with a value of three for such order (M = Nb and Tc) and the compound with formal bond order of two (M = Ru) from Rh.4.1.1 having one as formal bond order. However, for these specific cases the observed δ differences amount on average to about 0.3 and are therefore much lower than the difference of one between the corresponding formal bond orders (Table 9). One may thus safely conclude that the trends of the S%(M) and δ(M,M′) values almost agree also within this M_2(formamidinate)$_4$ series. Both trends concur with a largely enhanced covalent character (electron sharing) with respect to the two investigated 3*d* metal series and with a similar covalency for the compounds having formal bond orders of 3 and 4 or of 1 and 2 in the M_2(formamidinate)$_4$ series. Note that this apparently anomalous behavior, with respect to the formal bond orders, for both the δ(M,M′) and S%(M) values fits extremely well with the corresponding known anomalies [85] of the M–M bond lengths in the series.

Indeed the S(M) values exhibit an almost perfect linear correlation ($R^2 = 0.98$) with the M–M equilibrium distances through the series [14].

According to the molecular orbital theory, the decrease in the formal bond order in the M_2(formamidinate)$_4$ series past M = Mo is due to the progressive filling of the antibonding σ, π, and δ orbitals. Is this view reflected in the changes observed for δ(M,M′) and S%(M) through the series? And if so, how is it? The sum of $\delta(\Omega,\Omega)$ and half of $\delta(\Omega,\Omega')$ over all atoms Ω in a molecule yields its total number of electrons [38]. One finds that for Ω = M in the M_2(formamidinate)$_4$ series, the sum of δ(M,M) and of half the delocalization indices with all atoms linked to M, $0.5\cdot[\delta(M,M') + 4\cdot\delta(M,N)]$ recovers the M electron population within 0.2–0.3 electrons. The percentage of electron localization EL within the metal basin,[23] $EL\%(M) = [\delta(M,M)/N(M)]\cdot 100$, decreases from Nb to Mo [EL%(M) = 79.2 and 76.8] and then increases on going from Mo to Pd [EL(M)% = 76.8, 80.0, 84.1, 87.2, 90.4], in nice compliance, respectively, with the filling of the last available bonding MO in the Mo compound and with the progressive filling of the antibonding MOs past M=Mo. Hence, the number of localized electrons on the metal past Mo has to increase more rapidly than does its total electron population. The number of electrons shared by the metal with its four bonded N atoms, $0.5\cdot[4\cdot\delta(M,N)]$, shows instead only limited changes through the series, as it varies between 1.1 and 1.4 and it is so roughly independent from the nature of M. Combining the two results, it becomes evident that δ(M,M′) has to decrease through the series past Mo mainly because of a largely enhanced percentage of electrons localized on the metal, which is not compensated for neither by the almost stable electron sharing with the ligands nor by the continuous moderate decrease of the positive charge on the metal (Table 9). Progressive filling of antibonding orbitals through the series past Mo translates into an enhanced localization of electrons within the M basin and a decreased electron sharing between M and M′. The two processes lead to a general parallel decrease in the S%(M) values, although not in a such clear way as shown by the trends of decreasing δ(M,M′) and increasing EL%(M) values.

The 4*d* M_2(formamidinate)$_4$ complexes were previously studied by Llusar et al. [85] using an ELF approach. It is interesting to explore whether the ELF description of the M–M bond in such compounds has some point of contact with that provided by the SF analysis. It was found that the most important ELF topological feature characterizing the M–M bond in this series is the abnormally high value for the M–M core covariance, denoting that the fluctuations of the core basin populations of the two M are highly correlated one to another. A second, although less important ELF topological feature in these complexes is the occurrence of disynaptic metal valence basins. They are, in fact, characterized by extremely low populations and are even missing for M = Tc (besides M = Pd). Indeed, the population of the M core basins embodies more than 98% of the QTAIM metal electron population for

[23]The percentage of localized electron is computed with respect to the M pseudopotential valence electron population (in the case of the isolated atom, this is 14 electron for Mo, for instance).

all compounds in the series, with metal *d*-electron being almost completely included within these basins. The *d*-electrons are thus largely involved in the electron delocalization between ELF cores and considered as those electrons mainly responsible for the M–M interaction [85]. The somewhat unusual picture of bonding seems not at odds from that offered by the SF.

The net positive SF contribution from the metal to the density at the M–M bcp is the result of large positive LS contributions from atomic regions lying far apart from the bcp region, and not overriden by the smaller negative contributions arising from the region of positive Laplacian around the bcp. The smaller the distance from the bcp of the region of positive LS contributions, the larger is the expected SF positive contribution from the metal to the density at bcp. Analysis of the LS profiles along the M–M bond path with the bcp taken as rp indeed shows that the distance from bcp of the maxima of positive LS shows a nice inverse correlation with the S% values [14]. The more the *N* shell charge concentration region of the metal expands toward the bcp, the larger is expected to be the correlation of the electron fluctuations between the ELF core basins. Note that Llusar et al. [85] found an excellent correlation between the M–M distances and the core basin covariances, analogously to what we recovered between these same distances and the S(M) contributions at the bcp density.

The local bonding properties for the M–M bond in the M_2(formamidinate)$_4$ series are listed in Table 9. Examination of such properties shows once more unexpected trends versus the formal bond order. A positive $\nabla^2\rho_b$ is found for each M, with the lowest value for M=Ru and with the Pd compound, with formal bond order of zero and lowest δ(M,M′) and S%(M) values, having $\nabla^2\rho_b$ value about twice as large as that for M=Ru and M=Rh. It is thus clear that one cannot relate this indicator to the extent of electron sharing in the examined series, even if the dichotomous classification based on the sign of $\nabla^2\rho_b$ is taken with a reversed scale – the more positive the $\nabla^2\rho_b$ value, the greater the M–M bond strength – bearing in mind the missing outermost valence shell regions in the atomic Laplacian distributions of each M. Conversely, the $|V_b|/G_b$ ratio listed in Table 9 would describe all systems as featuring a M–M interaction which lies in the transit region for incipient covalent bond formation, even for the complexes with formal bond order of three and four and in spite of their very high δ(M,M′) and S%(M) values. On top of this, the highest $|V_b|/G_b$ ratio of 1.6 is not found for the M = Mo compound which has the highest δ(M,M′) and formal bond order values, but it is observed for the M = Ru and Rh complexes having δ(M,M′) values only one third as large and formal bond orders of 2 and 1, respectively. Finally, the ρ_b and the H_b descriptors – and less so the covalence degree H_b/ρ_b – roughly follow the trend of δ(M,M′) and S(M) values. Taken as a whole, one may conclude that a special care should be used when relating the 4*d* M–M bond nature to the (trend of) values of the local descriptors listed in Table 9, especially so if classification schemes based on the experience gained on bonding between atoms of the first three rows are adopted. A similar warning was evinced from the study of the saturated and, in particular, of the unsaturated binuclear 3*d* metal carbonyls. Other results common to such studies are recapped below.

The empirical relationships observed among δ(M,M′) and the S% contributions from the bonded metals M to the electron density at their intervening bcp have a particular relevance in view of the prominent role that the localization/delocalization index descriptors play in characterizing the M–M bonding in organometallic materials [66]. The possibility to get S% values directly from experiment adds an obvious importance to such a parallelism.

Even when the M–M bond is lacking and the M–M midpoint is used as an rp for evaluating the SF contributions, low δ(M, M′) values seem to come along with low S%(M) magnitudes. Implications of such an observation for the mechanism of indirect M–M bonding through 3c–2e interactions with the bridging ligands have been pointed out. In spite of not too unlike S%(M) magnitudes for bonded and unbonded M–M systems, the occurrence or lack of a M–M bcp is, however, clearly signaled by the quite distinct features of the LS profiles along the M–M internuclear axis in the two cases.

Four very recent studies [77, 80–82] have applied for the first time the SF analysis to the experimental electron densities of a number of binuclear metal complexes. A comparison with SF data derived from theoretical approaches was also provided in a number of cases. The main outcomes from such studies are briefly reviewed below, as for the insight drawn from the SF tool.

Overgaard et al. [80] carried out accurate diffraction experiments on a binuclear Co complex, composed of two $Co(CO)_3$ dimers bridged by an alkyne bearing one hydrogen and one cyclohexanol group. The charge density of the complex, which crystallizes with two molecules in the asymmetric unit, was determined using multipolar refinement of single-crystal X-ray diffraction data collected either with a synchrotron source at very low T (25 K) or using a conventional source at an intermediate T (100 K). The study thus offers several interesting potential opportunities. It allows for the comparison of the SF description in identical molecules using a single data set or the comparison of the SF description for each of the two molecules in the asymmetric unit using two data sets, obtained from different X-ray sources and collection temperatures. In addition, ab initio complete active space and DFT gas-phase calculations performed on one of the two molecules in the asymmetric unit permit a comparison between theory and experiment. Both agree as for the lack of a Co–Co bcp in the complex, whereas it proved impossible [80, 104] to locate in the experimental density[24] all of the four expected and theoretically found Co–C(alkyne) interactions. The S%(Co) contributions to the density at the Co–Co mp were small for both molecules and for all examined densities and in line with the values found for the binuclear Co complexes with similar Co–Co distances [14]. The S%(Co) values for the two molecules, whose structures may almost be perfectly overlaid and differ only in the relative orientations of the cyclohexanol moiety, are very much alike and slightly negative (about −18%/−16%) in the case

[24]It was indeed demonstrated that the extreme flatness of ρ in the CoC_2 triangles prevents the consistent location of the expected ring and bond CP in this region because their hypothetical experimental density differences would be surely much smaller than the typical multipole model residual density in the area.

of the experimental densities. The S%(Co) values are even smaller in magnitude and positive at about 4% for the single molecule studied theoretically. Interestingly, some of the atoms belonging to the other molecule in the asymmetric unit yield a small but nonnegligible contribution to the experimental density at the Co–Co mp, which may be one of the reasons for the observed S%(Co) difference respect to the theoretical value (the interaction with the other molecule in the asymmetric unit and in general the crystal field effects are clearly not taken into account in the gas-phase ab initio computation). Conversely, a comparison between the SF values obtained from the two experimental data sets would have enabled one to get a rough idea of how stable are the SF values with respect to the change of collected reflections and of the multipole model parameters refined thereof. Unfortunately, as for the scope of the present review, such a comparison was not reported in the paper, whereas one would argue that it could have been interesting in its own, in view of the assessment of the intrinsic stability of the SF description against those changes. One of the main reason that led Overgaard et al. [80] to analyze only the topological properties from the synchrotron data set, starting from page 3839 onward of their paper, was the closer match with theory observed for the density and Laplacian profiles along the various C–O bonds of the molecule, and in particular close to the bcp. Whether an assessment of the relative quality of two data sets against theory must be based on a comparison among local rather than integral quantities like the S% values is clearly debatable. Especially so, if one considers that local quantities are often too sensitive to otherwise small changes in the model approaches or data. An analysis of the different sensitivity of local and integral topological properties such as the SF values to moderate changes in the data and/or in their multipole refinement is briefly discussed in Sect. 4, using synthetic data from ab initio calculations [15].

Overgaard et al. examined the profiles of the LS along the Co–Co internuclear axis which showed for all densities and for both molecules a drop around the mp similar to that observed for the double-bridged $Co_2(CO)_8$ complex and comparatively missing in the M–M bonded unbridged form of the complex (see earlier). Such a similarity of the LS profiles in systems which lack a Co–Co bcp was taken by the authors as a clear corroborating sign of the absence of a direct Co–Co interaction in their investigated complex. Although probably correct, an "internal" comparison with a reference system having a Co–Co bcp is required to support this belief. The observed analogy with the results of the study by Gatti and Lasi [14] is also not stringent enough, since, as discussed earlier, the shapes of the profiles around the LS singularity can be safely compared among each other only when traced within the same distance from the singularity.[25]

Another interesting example in this area is due to Farrugia et al. [81]. These authors applied the SF, among other tools, to the study of a series of three binuclear Cu(II) coordination complexes $[Cu_2(ap)_2(L)_2]$, (ap = 3-aminopropanolate, L = nitrite, nitrate, and formate), using both experimental and theoretical electron densities. The complexes contain the same centrosymmetric alxoxy-bridged

[25]Information about the value of such a distance is not reported in [80].

motif, where each strongly Jahn-Teller distorted Cu(II) ion is ligated in its primary coordination sphere to three O atoms and one N atom in a square planar arrangement and with Cu–X (X = O, N) distances all below 2.05 Å. One axial coordination site is occupied by an oxygen atom of a pendant L^- anion from an adjacent complex in the lattice, while the other coordination site could potentially be considered as occupied by an additional O atom from the L^- anion already O-bonded to Cu(II) in the primary coordination sphere. For this second axial site, the angle subtended at the Cu(II) center is, however, much smaller than 90°. The Cu···O axial interactions all lie in the 2.45–2.70 Å range, suggesting their weakly coordinating nature. Graphical representations of the S% contributions for the various bcps in the examined complexes visibly and markedly distinguish the strong interaction of the Cu(II) ion with the ligators in the first coordination sphere from the much weaker ones in the secondary sphere (see Fig. 9 of [81]). Those for the former interactions are fairly localized and denoted by similar contributions from the bonded Cu and X = N, O atoms, summing up to about 65–70% of the bcp density and typical of strong metal–ligand interactions. Instead, those in the secondary sphere involve quite delocalized sources which extend over large parts of the complex and with S% contribution from the Cu atom opposite in sign to that of all atoms of the weakly linked ligand. The strikingly different portrait of sources in the two coordination spheres denote the shared nature of the metal–ligand interactions in the former and their largely polarized, non-shared nature in the latter.

The two Cu(II) centers, lying about 3 Å far apart, are known to be strongly antiferromagnetically coupled [81] but the lack of a Cu···Cu bcp and bond path suggests the absence of any direct metal–metal interaction in the complex. The molecular graph rather provides an experimental support for an exchange pathway via the bridging O-atoms, as also clearly suggested by the plot of the spin density in the plane of the $Cu_2(\mu\text{–}O)_2$ unit. When the rp is taken at the Cu···Cu midpoint, that is at the rcp of this unit, the S% contributions appear fairly delocalized, with the four atoms in the ring determining about 40% of the density at the rcp and with the largest contributions being those from each Cu atom, amounting to about 13%. The small sources from the metal atoms agree with the insignificant value of 0.02 found for the δ(Cu,Cu′) delocalization index and with the lack of a shared direct interaction between the two metal atoms. The δ(Cu, X; X = N, O) values are instead close to 0.5, corroborating the large difference observed for the SF portraits of the Cu···Cu and Cu–X interactions.

In their work on the dinuclear borylene complex $[\{Cp(CO)_2Mn\}_2(\mu\text{-B}t\text{Bu})]$ (Cp = C_5H_5), Flierler et al. [77] have applied, among other tools, also the SF analysis to discuss the nature of the Mn–Mn bonding in a system that represents a model compound for such kind of investigations in bridged and nonbridged organometallic complexes. The 18-electron rule would predict a bond between the two Mn atoms, as would be suggested also by their short internuclear distance of 2.78 Å and by the lack of any evidence for unpaired electrons in the complex. However, no bond path was recovered between the two Mn atoms, neither using the electron density derived from the low-temperature high-resolution X-ray diffraction

experiment nor that from a gas-phase ab initio calculation. A dominance of delocalized bonding via the bridging ligand was thereby suggested [77].

The S% values comply with a delocalized picture of sources when the Mn–Mn mp is taken as rp. The two Mn atoms are found to determine a small negative contribution to the electron density at the Mn–Mn mp, with S% from each Mn atom being about −8.5%. More than 50% of the density at the mp is instead determined by the carbonyl groups. The B atom also serves as a sink (S%(B) = −11.6%), as likely the result of the polarization of its Laplacian density due to bonding to the more electronegative C atom.[26] One anticipates, however, a large and positive S% contribution from the global B*t*Bu ligand, analogously to that found for the bridging carbonyls in the dinuclear carbonyl complexes, though an exact estimate of such a cumulative contribution from the ligand cannot be exactly deduced from [77] (the expected small S% contribution from the two Cp groups is not quoted).

A comparison between the LS profiles in $Mn_2(CO)_{10}$ and [{$Cp(CO)_2Mn$}$_2$ (μ-B*t*Bu)], along the Mn–Mn internuclear axis and with the Mn–Mn mp taken as a rp, was also not reported in the paper, though it might provide interesting insights. One could likely observe a comparatively larger drop of the negative LS around the mp in the borylene complex. Indeed Table 2 of [77] reports a positive Laplacian value at the Mn–Mn mp for the $Mn_2(CO)_{10}$ compound which is halved with respect to the one in the borylene complex, and the Laplacian is found to become even slightly negative for $Mn_2(CO)_{10}$ using the theoretical density. Hence, analogously to the LS profiles comparison reported earlier for the unbridged and the double-bridged $Co_2(CO)_8$ compounds, a close examination of the mentioned Mn–Mn LS profiles would allow to provide further evidence for indirect metal–metal bonding in the borylene complex.[27]

A parallel paper by Goetz et al. [78], with some common authors to [77], has examined the theoretical electron densities in a number of gas-phase dinuclear manganese complexes. These include, among other, two borylene-bridged compounds, one of which coincides with that studied experimentally by Flierler et al. [77] in the crystalline state. Both papers raise the important question of whether the bridging boron ligand may be better described as a substituted borane or as a true bridging borylene with an electron lobe from the B atom being directed at the Mn–Mn mp. Goetz et al. [78] found that the preference between the two bonding situations shows a large dependence on the adopted theoretical level. As a function of the exact-exchange admixture in the DFT functional, the boron bridge may involve either delocalized 3c bonding across the bridge or bonding through two

[26]Indeed, in the dinuclear carbonyl complexes, the large S% contribution from the CO ligands to the M–M mp or bcp is due to the dominant SF contribution from the indirectly bonded but more electronegative O atom, rather than from the directly bonded, but relatively less electronegative, C atom that exhibits negligible or sometimes even negative S% values (see for instance, Fig. 8).

[27]Note that these LS profiles comparisons are meaningful only because the bridged and unbridged forms share similar M–M distances. If it were not so, comparison of the Laplacian values and of LS profiles should take into account that the atomic Laplacian is smoothly decreasing in value with increasing distance from the metal.

2c separate bonds. Exact-exchange (hybrid functionals) seems to clearly favor the former situation. A given bonding scheme is assigned by Goetz et al. [78] on the basis of the number of $-\nabla^2\rho$ maxima in the valence shell charge concentration (VSCC) [6] of B or by inspecting the synaptic character of the ELF valence basins; 3c bonding is defined through the presence of one charge concentration maximum directed toward the M–M axis and of only one trisynaptic valence ELF basin, while the 2c bonding scheme is typified by the occurrence of one $-\nabla^2\rho$ maximum along each Mn–B direction or by the presence of two additional disynaptic ELF basins, besides the trisynaptic one. Note that the "borylene" compound discussed by Flierler et al. [77] is rather classified as a substituted borane, on the basis of the VSCC portrait of the B atom in the experimental Laplacian distribution. It thus appears that the precise assignment to one of the two bonding situations is quite challenging, since even small changes in the DFT functional or in the choice of multipole model expansion may cause the shift from one bonding scheme to the other. This is the typical case where distinction between two alternative bonding descriptions is given in terms of local properties, which may possibly undergo abrupt changes if the examined systems are characterized by distributions close to bifurcation catastrophe points (in this case of the $-\nabla^2\rho$ and ELF scalar fields). We thus propose that more insight on the borylene/borane debate could perhaps come from comparing the profiles of the SF contribution from the B atom to the electron density along the B–Mn and the B–(Mn–Mn mp) lines for any pair of complexes which are classified as substituted borane and borylene, respectively, in terms of the number of their boron VSCC $-\nabla^2\rho$ maxima. Examination of such profiles should allow to establish and quantify how diverse are in reality the bonding situations in the two systems. If a complex is close to a bifurcation point in its Laplacian distribution, so that tiny changes in the DFT or multipole model may shift it from the "borane" to the "borylene" assignment, one easily anticipates that very small changes in both the Mn–B and B–(Mn–Mn mp) profiles will be manifest between the two kinds of bonding schemes. Conversely, if the $-\nabla^2\rho$ maximum (a) is (are) quite evident, both profiles should show clearly distinct features in the two bonding cases. Also informative (and quite less computationally expensive) could be the comparison of the LS profiles along the same Mn–B and B–(MN–Mn mp) lines and taking the Mn–B bcp and the Mn–Mn mp as reference points, respectively. SF studies of such type have, however, not yet been performed.

We conclude this section by mentioning a recent communication by Overgaard et al. [82] on an ongoing experimental electron density study of the Mg–Mg bonding character in Mg(I) dimer complexes, containing anion-stabilized Mg_2^{2+} entities. Discovery [105] of thermally stable and relatively strong (~ 45 Kcal mol^{-1}) Mg–Mg σ-bond interactions in these complexes is unprecedented and shows that the remarkable case of decamethyldizincocene [106], which possesses a Zn(I)–Zn(I) single bond, can also be extended to the group 2 metal Mg. In view of the novelty and impact of Mg–Mg bonding, Overgaard et al. [82] analyzed a number of topological features for such bond, which shows a corresponding bcp with very small ρ_b and $\nabla^2\rho_b$ values amounting to about 0.015 and 0.010 au.

Evaluation of the S% contributions from the Mg at the Mg–Mg bcp reveals that the two Mg atoms determine about 30% of the bcp density, which is a small amount if compared to that of a normal shared covalent interaction like C–C having a δ(C,C′) value close to one, but large if compared to M–M bonding in the 3*d* dinuclear metal carbonyls and of similar size as that found for the insaturated dinuclear Co carbonyl with formal bond order of two. A comparison with corresponding values for model systems featuring homonuclear bonding between neutral and variably charged third row atoms would be helpful to place on a relative scale the observed S%(Mg) value. Also important would be to know the δ(Mg,Mg′) from a gas-phase calculation of the dimer. Precious further insight on bonding could then be obtained by evaluation of the S%(Mg) contributions for a model density made by the sum of the electron densities of the two noninteracting monomers, placed as in Mg(I) dimer complex. Operating this way, the effect of Mg–Mg bonding formation could be singled out and evaluated as we did previously for $Mn_2(CO)_{10}$ [14].

Overgaard et al. also discuss the LS profile along the Mg–Mg bond path showing "an intermediate-sized drop near the bcp (Fig. S7 in the supporting information of [82])" and noting that this drop should disappear with the onset of chemical bonding. This observation is probably not particularly meaningful in its own, as it should be supplemented by an LS profile study on a comparative scale, for instance on the Mg(I) dimer complex and on the corresponding noninteracting dimer, as performed for $Mn_2(CO)_{10}$ [14]. The remark is not to be meant as a criticism, but just as a suggestion for further work on this important, newly discovered metal–metal interaction.

3.4.3 Metal–Ligand (M–L) Interactions in Single Metal π-Hydrocarbyl Complexes and in Metal–Silane σ-Complexes

This section deals with the application of the SF approach to the interesting cases of (a) the delocalized interactions between transition metal (TM) atoms and open or close-conjugated π-systems [72, 83], and (b) the nature of metal silane σ-bond interactions in complexes formed by η^2 coordination of a ligand Si–H bond to a TM center, to provide an insight into the mechanism of Si–H bond activation within these complexes [84]. As we will show, it is the fairly delocalized and nonclassical character of all such interactions that makes SF a quite attractive tool for their study.

The nature of the chemical bond between transition metals and π-hydrocarbyl ligands has been the subject of several MO investigations, and it is in general considered to be well understood within this specific model approach (see [83] and references therein). Conversely, the well-known fluxional mobility of ligands in such systems challenges the usual QTAIM description of bonding in terms of two-center interactions and of corresponding lines of maximum accumulation of density (bond paths) between bonded pair of nuclei. Indeed, the very low barrier commonly observed in the gas phase for the rotation of ligands makes the actual structure

adopted in the crystal phase by the π-hydrocarbyl complexes to depend on kinetic as well as thermodynamic factors [83, 107, 108]. For instance, in the case of metallocenes such as ferrocene, $(\eta^5C_5H_5)_2Fe$, examples of staggered, eclipsed, and in-between ring geometries may be found in the Cambridge Structural Database [109] for this complex included as a guest in the lattice. Even in the gas phase, the search for the ground state minimum geometry of metallocenes is often critically dependent on the level of computation – the adequate treatment of electron correlation being a stringent condition for their correct geometry determination [110]. Because of the characteristic potential energy surface flatness for the metal-(π-hydrocarbyl) interactions, fewer M–C bond paths than expected on the basis of the formal hapticity are normally recovered, and their actual number is neither usually predictable nor easily explainable [83].[28] It is thus not surprising that based on a number of theoretical studies on metal-(π-hydrocarbyl) complexes [111, 112], Richard Bader et al. were led to conclude that "the bonding of a metal atom to an unsaturated ring is not well represented in terms of a set of individual bond paths, but rather by a bonded cone of density…" so that such bonding "… is best viewed as involving an interaction with the delocalized density of the entire ring perimeter, a picture that is conceptually similar to that conventionally used to denote the interaction of a metal with an unsaturated ring in chemical structure diagrams" [111]. Does the SF description reflect such a delocalized picture of bonding in metal π-hydrocarbyl complexes? How does it relate with the bond path view or with the nature of the interatomic electron sharing indicated by the corresponding delocalization indices?

Through a sequence of two outstanding papers, Farrugia et al. [72, 83] have given illuminating answers to all these questions. In the earlier one, a combined experimental and theoretical charge density study of the trimethylenemethane (TMM) complex $(Fe(\eta^4\text{-}C\{CH_2\}_3)\text{–}(CO)_3)$ was performed [72]. The complex provides an example of a delocalized π-hydrocarbyl system where the unsaturated C–C bonds are arranged in a stellated fashion rather than in a ring or chain. Agreement between the theoretical and experimental topologies was excellent. Curiously enough, both density topologies found only one bond path between the TMM ligand and the Fe atom, from the C_α atom, whereas none of such paths was recovered between the metal and any of the three C_β atoms (Fig. 11a, b).

According to the bond path criterion, the complex should be described as a η^1-complex, but normal-mode analysis, ESCA and photoelectron spectra, NMR barrier to rotation of the TMM ligand in the complex, and qualitative analysis of the Kohn-Sham orbitals seem to all suggest a significant π-interaction with the C–C bonds ([72] and references therein). The available physicochemical evidences, save

[28] It was noted, however, that the virial graphs, at variance with the molecular graphs, do reflect in general the formal hapticity in metal π-hydrocarbyl complexes [83]. Virial graphs are the analogous of molecular graphs for ∇V replacing $\nabla\rho$ [6]. Because of their general homeomorphism, these two vector fields usually exhibit topologically equivalent graphs, but this may not hold true when the equilibrium structure is close to a catastrophe point for at least one of the two fields.

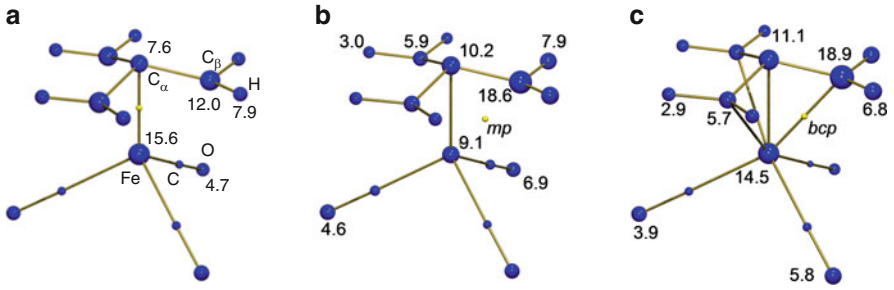


Fig. 11 Source function percentage contributions from atomic basins in the trimethylenemethane (TMM) complex (Fe(η^4–C{CH_2}$_3$)–$(CO)_3$) to the electron density at different reference points (denoted by a *small yellow sphere*). These are (**a**) the Fe–C_α bcp; (**b**) the Fe–C_β midpoint (mp); (**c**) the Fe–C_β bcp for a slightly deformed geometry of the complex, where the Fe–C_α–C_β angle is reduced from the optimized equilibrium value of 76.6° to 73°, while retaining the original C_{3v} symmetry. This small geometry change yields to the formation of three new bond paths linking iron to the C_β carbons also (adapted from Fig. 7 and Fig. S11, with permission from [72], Copyright 2006, American Chemical Society)

the molecular graph, would so imply a significant interaction of iron also with the C_β atoms and a consequent classification of the complex as a η^4 one.[29]

It was thus not surprising that a picture at odds with that provided by the bond path criterion emerged from the (Fe,C) delocalization indices and the SF analysis (both were only performed on the gas-phase theoretical densities). Instead of being negligible, the δ(Fe,C_β) value, 0.571, is large and about 40% larger than the δ(Fe, C_α) value, equal to 0.369. Similarly, the atomic SF contributions to the density at the Fe–C_α bcp exhibit a pattern which closely resembles that obtained when the rp is taken at the Fe–C_β midpoint (mp) (Fig. 11a, b). Even more, the percentage contribution from C_β to the Fe–C_α bcp density is larger than the corresponding one from C_α to the Fe–C_β mp density. In both cases the sources are fairly delocalized, with cumulative contributions from the three methylene groups being equal to about 60% of the total rp density, to be compared with a contribution from the central C atom of about 7–10% only. The sum of contributions from the iron atom and the bonded C_α atom is small and about 20% for both rp densities, while that from the iron and the pertinent C_β atom is also approximatively constant but slightly larger, amounting to about 27%. Not insignificant is also the contribution from the three carbonyl groups (about 20% for both rps). All these values imply a delocalized picture for the metal-(π-hydrocarbyl) interaction with a significant π-interaction with the C–C bonds and a clear involvement of the methylenic H atoms as well.

[29]The (TMM) complex, analogously to what found in other metal π-hydrocarbyl complexes, represents the uncommon case of the absence of a bond path where chemical intuition would expect a bond, which is an opposite situation to the more common recovery of many bond paths where chemists would not draw any bond line.

The experimental and the gas-phase ab initio geometries of (Fe(η^4-C$\{CH_2\}_3$)–$(CO)_3$) are close to structural instability; that is, new structures may occur even for small geometry changes. By decreasing the Fe–C_α–C_β angle from the optimized equilibrium value of 76.6° down to 73°, while retaining the original C_{3v} symmetry, three new bond paths linking iron to the C_β carbons appear [72]. Despite this evident structural change – namely from a η^1 to a η^4 complex, according to the bond path criterion – the (Fe,C_α) and the (Fe,C_β) delocalization indices are hardly changed, as they decrease to 0.351 and increase to 0.609, respectively. Analogously, no significant change is observed in the SF pattern when the rp is taken at the Fe–C_β bcp in the distorted geometry, rather than at the Fe–C_β mp in the minimum energy structure (Fig. 11c). Overall, due to the proximity of the optimized geometry to structural catastrophe, the SF contributions look almost the same, whether a Fe–C_β bcp is present or lacking. In both cases, the contributions from the C_β atoms to the density at the rp are greater than observed from the C_α atom. It is thus gratifying that the delocalization indices and the SF provide pictures of bonding in this complex, which nicely fit within each other and also with the other available experimental evidences. Clearly the occurrence or lack of a bond path is also a physical outcome, and the bond path criterion should not be ignored because of its apparently contrasting conclusions. Simply, the nonlocal descriptors seem more appropriate for describing an inherently nonlocal interaction such as that of TMs with the delocalized π-hydrocarbyl ligands, as pointed out earlier for other interactions characterized by flat energy surfaces and electron densities.

Similar conclusions were drawn by Farrugia et al. in their combined experimental and theoretical charge density study [83] of three different metallocenes (η^5-C_5H_5)$Mn(CO)_3$, (η^6-C_6H_6)$Cr(CO)_3$, and (*E*)-{(η^5-C_5H_4)CF=CF(η^5-C_5H_4)}(η^5-C_5H_5)$_2Fe_2$. For the sake of conciseness, we focus here only on the first of such compounds whose experimental molecular graph and SF percentage sources for a number of representative rps are shown in Fig. 12. The molecular graph (Fig. 12a) shows only four Mn–C bond paths, one less than expected on the basis of formal hapticity. Curiously enough, the bond path is missing for the shortest (Mn–C7) rather than for the longest Mn–C_{ring} distance.

The Mn–C5 bond path is rather curved, and its bcp is 0.02 Å close to the C4–C5–Mn rcp, indicating that these two CPs are proximal to coalesce to yield a structural evolution. Indeed, this bond path was even not recovered with some of the investigated multipole models, whereas the theoretical density for the gas-phase-optimized structure exhibits the number of bond paths expected from the formal hapticity of the complex [83]. Similar discrepancies were found also for the other studied systems, which illustrate once more the difficulties encountered when using a bond path approach to characterize the delocalized metal-(π-hydrocarbyl ligands) bonding.

The experimental SF patterns shown in Fig. 12b–f are instead very similar to those obtained from the ab initio density (Fig. 9 of [83]) and visibly distinguish the quite different types of chemical interactions present in the complex. The SF contributions to Mn–C(O) bcp (Fig. 12b) are typical for such interaction and very

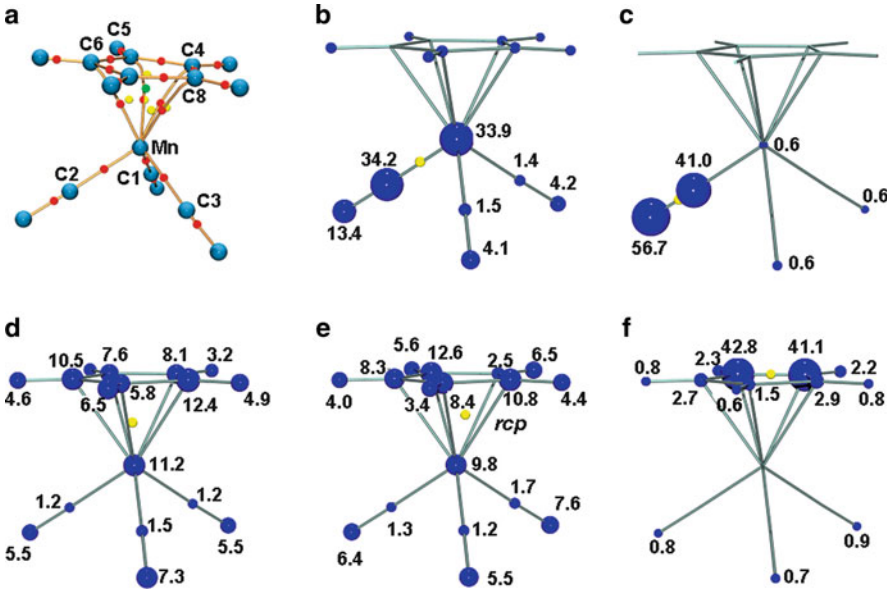


Fig. 12 Metal–ring and other chemical interactions in (η^5-C_5H_5)Mn(CO)$_3$ using the bond path and the SF percentage descriptors; (**a**) the experimental molecular graph (bcps are shown in *red*, rcps in *yellow*, and cage critical point in *green*); (**b**–**f**) source function percentage contributions from atomic basins to the experimental electron density at different reference points (denoted by a *small yellow* sphere). These are (**b**) the Mn–C_2 bcp; (**c**) the C_2–O bcp; (**d**) the Mn–C(Cp) bcp; (**e**) the Mn–C(Cp) rcp; (**f**) the C–C bcp. Absolute source contributions less than 0.5% are not shown (adapted from Figs. 2 and S23, with permission from [83], Copyright 2009, American Chemical Society)

much alike to those found for terminal M–C(O) bonds in the dinuclear metal carbonyls. The two bonded atoms yield about 68% of the bcp density and the carbonyl O atom an additional 13%. As shown by Farrugia et al. [72], the large contribution from the carbonyl O atom is not to be simply interpreted as a signal of Mn–CO π back-bonding, since similar percentage source contributions are found for the B–C bcp in the adduct $H_3B{\leftarrow}CO$ where no B–C π back-bonding is possible. In our view, this large SF contribution from the carbonyl O atom results from the nonnegligible portion of the 3σ HOMO lone pair orbital of CO lying into the O basin ([103] and page 10061 of [12]) and the dominant role that such orbital plays in the dative σ bond to the metal [103].

When the rp is placed at the C–O bcp, the global contribution from the two bonded atoms rises to about 98% (Fig. 12c) as found for all other metal-carbonyls thus far investigated. It is not just the sum which is transferable, but also the separate contribution to this sum from the C and the O atoms, equal to about 41% and 57%. The Mn–C_{ring} bcps and the rcps pertaining to the three-membered rings formed by the Mn and any two metal-bonded neighboring C atoms of the π-hydrocarbyl ligand lie on an annulus of almost constant electron density and

vanishing density gradient [83]. The SF contributions patterns nicely reflect such a delocalized nature of the Mn-ring interaction. All basins of the π-hydrocarbyl ligand determine in a significant way the electron density in the annulus, and relative contributions from all basins are almost unchanged regardless the rp is at a bcp (Fig. 12d) or at an rcp (Fig. 12e). As a consequence, the SF pattern for the Mn–C7 interaction, which lacks a bcp, was found to be almost indistinguishable from that of the remaining four bonded Mn–C_{ring} interactions, provided the corresponding rp is also taken to lie in the annulus. The SF contribution from the $Mn(CO)_3$ group is significant too and independent of the rp position within the annulus. It amounts to about 30%, one third of it coming from the Mn atom. Taken as a whole, the SF patterns for the metal–ring bonding suggest not only that such interaction is delocalized through the whole ring, but that it involves, besides the metal atom, also the carbonyl ligands bonded to the metal. Finally, Fig. 12f shows the S% pattern when the rp is taken at the $(C–C)_{ring}$ bcp. The two bonded C atoms determine about 84% of the density at their bcp, which is consistent with that found for benzene (84.7%, Table 4). Even contributions from the other C atoms in the ring and from the H atoms are very much alike to those obtained for benzene. According to Farrugia et al. [83], the dominant S% contribution of the two bonded C_{ring} atoms to their bcp density may be seen as surprising, given the delocalization of the π density in these conjugated rings. However, we did already show in Sect. 3.2.2 that the effect of π-conjugation is visible through the SF approach, though it is clearly dampened when analyzed by placing the rp in the nodal plane. In that case, only the indirect effect of π-conjugation on the σ-distribution becomes manifest, which is to a good approximation the situation being analyzed in Fig. 12f (the ring plane may be roughly considered as a "nodal" plane since the metal–ring interaction is a small perturbation with respect to the strong covalent interactions within the ring, as the likeness with the SF pattern found for benzene would confirm). In order to get further insight on the perturbing effect of the metal–ring interaction on the electron distribution of the hydrocarbyl ring (and in particular that of "π-electrons"), we suggest to compare the SF patterns when the rp is displaced above and below the $(C–C)_{ring}$ bcp as we did for benzene. Differences in S% patterns when the rp is closer or farther to the metal atom would allow to estimate the perturbation of the π-electron conjugation and distribution which is caused by the onset of the metal–ring interaction.

Discussion in [83] is supplemented with the SF analysis of the theoretical densities of a homologous series of 18-electron model compounds having the formula $(\eta^n C_nH_n)M(CO)_3$ and varying ring sizes which range from $n = 3$ to $n = 8$, according to the change of the metal from Co($n = 3$) down to Ti ($n = 8$) along the first TM series. Also for these model compounds, an unpredictable number of metal–C_{ring} bond paths are observed, and fewer M–C bond paths than expected from the formal hapticity are found, for ring sizes greater than four. However, both delocalization indices and S% patterns agree in describing a similar level of bonding for all M–C_{ring} interactions, regardless of the presence or lack of a corresponding bond path and in full compliance with the expected chemical picture. The reader is addressed to the original paper for more details.

Farrugia et al. [83] also investigated a number of cases where the geometrical distortion in the metal–η^n-hydrocarbyl interaction is sufficiently large to yield a formal reduction in the hapticity. In such cases, it was pleasingly found that both the delocalization indices and the SF patterns clearly discriminate the formally bonded from the formally nonbonded M–C_{ring} interactions.

One minor point remains to be clarified concerning this otherwise excellent paper by Farrugia et al. [83]. When discussing the (*E*)-{(η^5-C_5H_4)CF=CF(η^5-C_5H_4)}(η^5-C_5H_5)$_2$Fe$_2$ system, it is mentioned that the weak C–H···F interaction provides an interesting case, with "characteristics and highly delocalized" sources and with the three interacting basins all acting as very large "sinks" for the density, while the other F atom provides the single largest source. We are not interested here in the discussion of this specific SF pattern, but concerned with the physical interpretation that was given for atoms acting as "sinks." Farrugia et al. indeed note that "when a basin acts as a sink, the kinetic energy dominates over the potential energy density when averaged over that basin." Clearly this is not true, since for the very definition of QTAIM basins, the kinetic energy density and the potential energy densities, when averaged over Ω, fulfill the atomic virial theorem, and the ratio of their averaged magnitudes has to be always equal at equilibrium to the virial ratio of 1:2 as it is for the global system. As shown in (5), an atom yields a negative source if LG dominates over LV when averaged over its basin. Although LG and LV are related to G and V, they clearly differ from them because of their Green's or influence function term (Sect. 2.1). It is this term which causes an atom to act as a source or as a sink, depending on the position of the rp, as shown earlier in this chapter (Sect. 3.3) for the H atom involved in weak or moderate strength OH···O bonds.

The chemistry of σ-bond complexes formed by η^2 coordination of a ligand *E*–H bond (*E* = C, Si, H, B, Sn or Ge) to a TM center has been the subject of intense interest over the past three decades, as these systems provide an insight into the *E*–H activation of *E*–H bonds by TM centers [113]. Silane σ-bond complexes, which were the first to be isolated and recognized back in 1969 [114], presently represent the second largest class of σ-bond complexes after molecular hydrogen systems and serve also as a model for their more ephemeral alkane σ-bond cousins and for C–H activation [115]. Recently, Mc Grady et al. [84] have investigated the nature of metal silane σ-bond interaction in a number of key systems by a range of experimental and computational techniques, including an SF analysis, which is briefly reviewed below. In particular, their study focused on three simple Schubert-type [116] complexes [Cp′Mn(CO)$_2$(η^2-HSiXY$_2$)], with *X* = H, F, and Cl, respectively, and *Y* being Ph for *X* = H, F or being Cl for *X* = Cl (the molecular scheme and structure for *X* = F, *Y* = Ph is shown in Fig. 13a, b). This series of complexes may be viewed as "snapshots" at various stages along the reaction coordinate for oxidative addition of the Si–H bond to the metal, although the determination of the actual reaction stage in a given complex turns out to be difficult because X-ray diffraction fails to locate the H atoms with sufficient accuracy. Earlier topological analyses of the charge density [111] and photoelectron spectra ([117] and note 17 of [84]) have indeed classified the compound with *X* = *Y* = Cl as a nearly complete

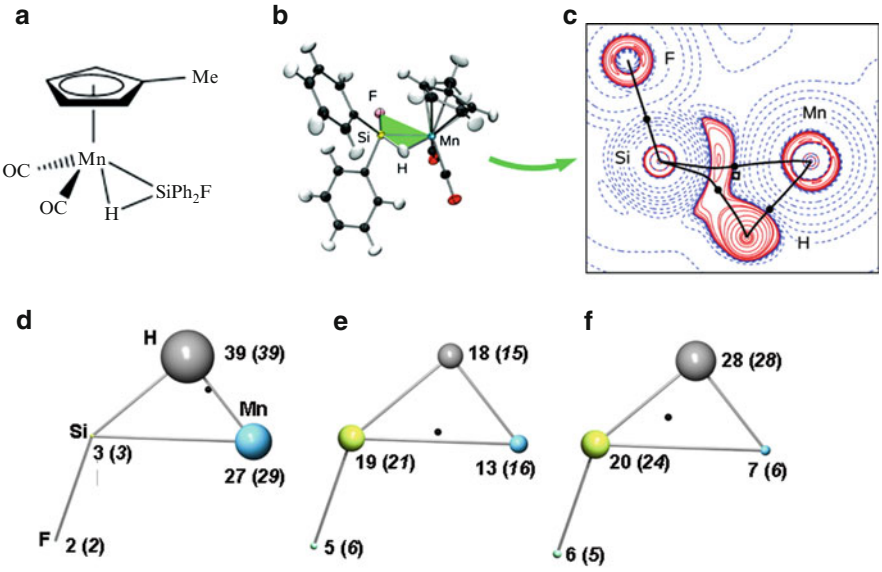


Fig. 13 Nature of the bonding in the metal–silane σ-complex [Cp′Mn(CO)$_2$(η^2-HSiFPh$_2$)] (Cp′ = η^5-C$_5$H$_4$Me); (**a**) molecular scheme; (**b**) molecular structure; (**c**) $\nabla^2\rho(\mathbf{r})$ contour map in the Mn–H–Si plane. Negative and positive values are marked by *solid* and *dashed lines*, respectively. Bcps and Rcps are marked by *closed circles* and *squares*, respectively; the bond paths are shown by *solid lines*; (**d–f**) source function percentage contributions from atomic basins to the electron density at different reference points (denoted by a *small black sphere*). These are (**d**) the Mn–H bcp; (**e**) the Mn–Si bcp; (**f**) the Si–H bcp. The reported values are derived from the experimental and from the DFT (in *parenthesis* and in *italic*) electron density distributions (adapted from Scheme 1, Fig. 7, and Table of Contents, with permission from [84], Copyright 2009, American Chemical Society)

oxidative silane addition product, with only a negligible residual Si–H interaction, while the other two complexes with X = H or F were identified as silane σ-bond complexes at an early stage of Si–H bond addition [118]. In contrast to these earlier reports, Mc Grady et al. [84] found that the Mn–Si–H bonding in the last two complexes is not fundamentally different from that in the $X = Y$ = Cl compound which possesses more than one electronegative ligand at Si. In spite of a shorter Mn–Si bond, the Si–H and in particular the Mn–H interactions appear not significantly weaker than for the X = H or F compounds. Mc Grady et al. [84] could arrive to these conclusions on the basis of their neutron diffraction study of the complex with X = H,[30] the available neutron structure for the complex with X = F [119], and careful X-ray diffraction and DFT charge density studies for all the three complexes. Furthermore, inspection of the MOs of each complex and of the frontier

[30]Note that this is the only example of a neutron structure for a complex containing both a coordinated and an uncoordinated Si–H moiety in the vicinity of the metal atom.

orbitals of their silane and metal fragment moieties led Mc Grady et al. [84] to formulate a new and more sophisticated Mn (η^2-SiH) bonding model, improving and replacing the previously established σ(Mn–H) → σ*(Si–*X*) mechanism to explain the electronic influence of the auxiliary *X* ligand *trans* to the activated Si–H bond. It was found that the Mn (η^2-SiH) bonding in all of these systems occurs through an asymmetric oxidative addition reaction coordinate in which the Mn–H bond is formed at an early stage, whereas Mn–Si bonding is being controlled and enforced by the extent of Mn → σ*(*X*–Si–H) π-back donation. This electron flow from a filled metal orbital of proper symmetry into a three-center ligand orbital displaying both Si–*X* and Si–H antibonding character results in a simultaneous activation of both the η^2-coordinating Si–H bond and the Si–*X* bond in *trans* position. As a consequence, the control role exerted by the *trans*-oriented *X* ligand on the geometry of the Mn(η^2-SiH) moiety becomes manifest: the higher the electron-withdrawing character of *X*, the greater the Si–*X* and Si–H bond activation, owing to the increased Mn → ligand π-back donation. This MO model interpretation (clearly available only at a theoretical level) was, however, unequivocally corroborated by an SF analysis on both the experimental and theoretical charge densities of the three investigated complexes. Figure 13d–f displays S% contributions from atomic basins to the electron density at different bcps of the Mn(η^2-$HSiFPh_2$) moiety. When the rp is at the Mn–H bcp (Fig. 13d), the dominant contributions come from the Mn atom (27%; *29%*, experimental values in *italic*) and the H atom (39%; *39%*), whereas the contribution of Si is indeed small (3%; *3%*) as expected for a strongly localized Mn–H bond. Note also that almost unchanged percentage contributions are found for the other two studied complexes. This result confirms that the oxidative addition of the silane ligand to Mn is an asymmetric process in which the Mn–H is formed at an early stage and that such bond, as expected, cannot be significantly affected by the extent of the Mn → σ*(*X*–Si–H) π-back donation. The situation appears quite different when the rp is placed at the Mn–Si bcp (Fig. 13e). Here Mn, Si, and H each contribute to a very similar extent to the bcp density (13%, 19%, 18%; *16%*, *21%*, *15%*, for Mn, Si, and H, respectively, and with experimental values in *italic*). Such delocalized sources imply a strongly delocalized interaction, with formation of the Mn–Si bond affecting directly the Si–H bond (hence the H atom percentage contribution) because of the π-back donation from the metal into the antibonding σ*(*X*–Si–H) orbital. Analogously, when the rp is placed at the Si–H bcp (Fig. 13f), very delocalized sources are again observed, with a nonnegligible contribution from the Mn atom. According to the Mn → σ*(*X*–Si–H) model, the S%(Mn) was found to theoretically increase from $X = H$ (5%), to $X = F$ (6%) and to $X = Cl$ (9%), confirming that the back-donation from Mn to Si increases as the Si center becomes more positive.

The SF analysis "translates" to a set of percentage atomic contributions the information in Fig. 13c, where the Laplacian distribution of the density in the Mn–H–Si–F plane, overlaid with the bond paths within the Mn(η^2-HSiF) moiety, is displayed. It is gratifying to see that the SF provides not only a precious chemical insight into an intriguing bonding scenario, but also a physical validation, both at a theoretical and at an experimental level, of the MO model interpretation of bonding

in Schubert-type complexes. Si–Mn bonding, which dictates the strength of the Mn (η^2-SiH) interaction, is found to be the result of a complex interplay of contributions from all four atoms of the Mn(η^2-HSiX) moiety.

4 Is the Source Function a "Robust" Descriptor?

Local descriptors derived from the topological analysis of various scalar fields – including the electron density as the most relevant one – play a central role in the majority of the existing schemes for classifying chemical interactions in gas phase and in crystals [2]. Use of local descriptors is simple, immediate, and computationally inexpensive, but it has the drawback of emphasizing the role of the properties of the scalar field under examination at the few selected points (e.g., bcps) only. Thus, the approach not only assumes that these points be the most representative for the entire bonding interaction – which instead in general involves a not too localized rearrangement of the scalar field distribution with respect to the sum of the atomic distributions – but also suffers from the usually high sensitivity of a local property to changes in the way the associated scalar field is derived. For instance, if the dichotomous classification [97] based on the sign of $\nabla^2\rho$ is used, with shared and closed-shell type interactions being characterized in terms of negative and positive $\nabla^2\rho_b$ values, a very large dependence on the model used to build $\rho(\mathbf{r})$ may be observed on $\nabla^2\rho_b$. The Laplacian at bcp is given by the sum of the three curvatures of the density at this point, and significant changes in these curvatures, in particular in the one parallel to the bond path, are found to occur when different multipole models, or Hamiltonian and basis sets are adopted for experimental and theoretical electron densities, respectively [2, 32]. The electron density difference between an ab initio density and the one derived from multipole modeling of the former density projected onto structure factors allows for a direct estimate of the bias introduced by the multipole model (MM) refinement step. Hereinafter in this section, the original density and that derived from it through the MM will be referred to as the primary density (PD) and the multipole-modeled primary density (MMPD), respectively. In the unattainable limit of zero bias, the two densities will be equivalent, but in practice their difference may often be relevant. Differences in their associated $\nabla\rho$ and $\nabla^2\rho$ fields, which define the position of the bcps and the $\nabla^2\rho_b$ values, respectively, may be *at fortiori* very significant and were indeed found unacceptably large in several cases [15].

The less will be affected by changes in the way $\rho(\mathbf{r})$ is obtained, the more robust and informative will be a given local bond descriptor, provided these changes lead to otherwise physically similar electron densities and with similar general accuracy. Changes in $\rho(\mathbf{r})$ and in its derivatives will clearly influence also the SF and to an extent dependent on the way it is being analyzed (i.e., in terms of atomic, percentage atomic, or local contributions). When comparing SF results from PD and MMPD densities or from any two different $\rho(\mathbf{r})$ distributions, three major interrelated sources of discrepancies may come to the play – the change of $\rho(\mathbf{r})$

within an atomic basin, the variation of its basin boundaries, and the shift in the bcp position.

The robustness of the SF descriptor against changes in the way $\rho(\mathbf{r})$ is obtained and/or against the bias introduced by multipole modeling on a PD has been investigated in a recent preliminary study by Lo Presti and Gatti [15]. Three test systems were considered: (a) a small organic molecule (hexafluorocyclobutene, C_4F_6); (b) carbon monoxide, CO, and (c) bis(pentacarbonylmanganese), $Mn_2(CO)_{10}$. In all three cases, the PD for a crystal made of noninteracting molecules and the corresponding MMPD density, using different MMs, were compared. In addition, diverse PDs were considered for CO to test the effect of the choice of the Hamiltonian and of the basis set, whereas for $Mn_2(CO)_{10}$ the multipole model experimental density (MMED) obtained from the MM refinement of a set of experimental structure factors [71] was also included in the comparison. The main results obtained are briefly reviewed below, while the reader is addressed to the original paper [15] for more technical details.

Table 10 reports ρ_b and $\nabla^2\rho_b$ values for C_4F_6 and CO at their various bcps (see Scheme 3 for C_4F_6 atomic labeling) and the corresponding atomic S% contributions to ρ_b.

Comparing PD and MMPD data for C_4F_6, one observes that the S% contributions are generally very much alike: the largest discrepancies concern the polar C–F bonds, but even for these bonds the percentage differences never exceed 6 points, and for both densities the F atom yields a much larger density contribution at bcp than does its bonded partner. Conversely, differences in the ρ_b values may be as big as 24% with the largest ones being, as expected, for the C–F bonds. Discrepancies between PD and MMPD data become impressive and hardly acceptable when the $\nabla^2\rho_b$ values are compared. Deviations as large as 600% in magnitude are observed for C–F bonds, for which the PD Laplacian values are far less negative than following the MM treatment. In the case of the F1–C2 interaction, the MM bias even leads to a change in the $\nabla^2\rho_b$ sign, notwithstanding an appreciably positive PD $\nabla^2\rho_b$ value. It is remarkable that the S% contributions from C and F to the F–C bcp densities differ by a maximum of 6 points[31] only after the MM refinement of theoretical data, given that these contributions are evaluated in terms of the Laplacian distributions which show such impressive discrepancies at the C–F bcps.

Before analyzing why the S% appears as a much more robust descriptor than it is, in particular, $\nabla^2\rho_b$, we comment the results obtained on CO. The electric dipole moment for this molecule is known to be quite small, despite the significantly high net atomic charges. The capability of reproducing the experimental dipole moment, which amounts to 0.11 Debye [123] and is directed from carbon to oxygen $^-C{\equiv}O^+$, is known to be very sensitive to the employed theoretical model (see for instance

[31]As explained in detail at page 311 of [15], changes in S% values are given as percentage point changes, rather than as percentage changes as done for ρ_b and $\nabla^2\rho_b$, because the S% values are already defined on a relative basis.

Table 10 Comparison of *X*–*Y* bond critical properties and atomic source function percentage contributions at *X*–*Y* bcp in C_4F_6 and in CO as derived from a theoretical primary density (PD) and its corresponding multipole-modeled primary density (MMPD)[a,b]

System/ model[c]	*X*–*Y*	*R*, Å	μ_z(D)[d]	ρ_b (e Å^{-3})	$\nabla^2\rho_b$ (e Å^{-5})	S%(*X*)	S%(*Y*)	S%(*X* + *Y*)
C_4F_6	F2–C1	1.327	–	1.79	−3.8	45.5	35.1	80.6
				2.22	*−31.3*	*51.7*	*33.7*	*85.4*
				(+23.9%)	(+724%)	(+6.2)	(−1.4)	(+4.8)
	C1–C1′	1.543	–	1.70	−14.7	34.5	–	69.0
				1.86	*−16.8*	*35.1*	–	*70.2*
				(+9.4%)	(+14.2%)	(+0.6)	–	(+1.2)
	C2–C2′	1.315	–	2.33	−23.8	41.8	–	83.6
				2.51	*−26.6*	*42.2*	–	*84.4*
				(+7.7%)	(+11.8%)	(+0.4)	–	(+0.8)
	C1–C2	1.493	–	1.83	−16.3	35.3	38.9	74.2
				1.93	*−17.2*	*34.9*	*39.9*	*74.8*
				(+5.5%)	(5.5%)	(−0.4)	(+1.0)	(+0.6)
	F1–C2	1.298	–	1.85	+6.2	47.6	39.1	86.7
				2.27	*−29.0*	*52.7*	*37.2*	*89.9*
				(+22.7%)	(−568%)	(+5.1)	(−1.9)	(+3.2)
CO	C–O	1.136	−0.134	3.36	−3.8	42.2	57.8	–
BVWN[e]/ Valenzano[f]				*3.47*	*−24.2*	*44.2*	*55.8*	–
				(3.3%)	(+537%)	(+2.0)	(−2.0)	–
CO		1.146	+0.237	3.06	39.1	40.9	59.1	–
BVWN[e]				*3.30*	*−32.2*	*46.1*	*53.9*	–
/cc-pVDZ[g]				(7.9%)	(−182.4%)	(+5.2)	(−5.2)	–
CO		1.110	−0.146	3.35	53.2	38.9	61.1	–
HF/cc-pVDZ[g]				*3.40*	*9.8*	*40.7*	*59.3*	–
				(1.5%)	(−81.7%)	(+1.8)	(−1.8)	–

[a]Data from [15]

[b]For bond topological properties and S% contributions, the first row refers to PD and the second row (italic) to MMPD data. The third row lists the percentage difference [(MMPD-PD)/PD]*100 between MMPD and PD values in the case of the bond topological properties, while for the SF percentages, it reports the percentage points difference between the PD and MMPD S% values

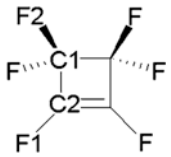
[c]In the case of the CO system, different combinations of Hamiltonian and basis sets have been investigated (see text)

[d]μ_z is the dipole moment component along the internuclear axis. It is taken as positive if the C atom is negatively charged

[e]BVWN is the DFT functional composed by Becke's exchange functional [34] and Wosko Wilk and Nusair correlation functional [120]

[f][121]

[g][122]



Scheme 3 Atomic labeling in C_4F_6

[124]). The bcp properties of CO, for either the isolated molecule or the molecule acting as a ligand, depend even more strongly on the chosen computational scheme or MM refinement of structure factors [2, 6, 125]. The $\nabla^2\rho_b$ value for CO is indeed extremely sensitive to the location of the bcp, which most often lies close to the nodal $\nabla^2\rho$ surface separating the core-shell charge depletion region of the carbon from the single, merged VSCC of CO. Despite numerous evidences of a shared interaction character of the C–O bond [125], $\nabla^2\rho_b$ is found often positive, but even very small environmental or modeling changes, capable to slightly shift the bcp toward the oxygen atom, may lead to negative $\nabla^2\rho_b$ values [2, 66]. The three PD models listed in Table 10 for CO differ as for their minimum energy geometry and dipole moment estimate. The corresponding S% results are rather stable, with differences in values among the three models not exceeding 3 points, whereas ρ_b and in particular $\nabla^2\rho_b$ show a much larger variability. The electron density values at the C–O bcp differ among each other by a maximum of about 10%, but the largest $|\nabla^2\rho_b|$ value is 14 times higher in magnitude and with sign reversed than the lowest one. When PDs data are compared with the corresponding MMPD results, one observes that the bias due to the multipole model refinement leads to acceptably limited changes on the ρ_b and S% values, the largest ones being about 7.9% and 5 points, respectively. Conversely, the $\nabla^2\rho_b$ values are not at all reproduced, with changes even as large as 530% in one case and with a value qualitatively similar in magnitude, but reversed in sign in another case. Table 10 reports MMPD data for the MM with refined k and k' screening parameters [10]. Adoption of standard unitary screening parameters, though leading to quite different numerical values for $\nabla^2\rho_b$, does not qualitatively affect the picture [15]. It is worth noting that, despite the large variations observed for the $\nabla^2\rho_b$ magnitudes with changes in the PD models or following the MM refinement of their corresponding structure factors, the S% values remain reasonably stable for such shared and partially polar bond. This holds true even when the $\nabla^2\rho_b$ sign changes. The large and positive QTAIM net charge of C and the location of the bcp closer to the C than to the O nucleus easily explain why the S%(O) always exceeds that from C.

The results obtained for C_4F_6 and CO seem to indicate that the S% values are carrying a more robust and thus more chemically meaningful information than do the ρ_b and, in particular, the $\nabla^2\rho_b$ local descriptors. Indeed, being the result of integration over an atomic basin, the SF or the S% contribution averages out the local bias introduced by the MM refinement or by the PD model change on the LS integrand (which is defined through the Laplacian). One also observes that since the SF contributions simply reconstruct the electron density, PD and MMPD SF results should roughly differ as their electron densities do. However, rather then being related to the ρ_b values, the S% contributions express how ρ_b is comparatively shared between the two linked atoms. Because of the dependence on the relative contribution to ρ_b rather than to the ρ_b magnitude itself, the S% descriptor turns out to be generally more stable than it is ρ_b (and a fortiori $\nabla^2\rho_b$) against both the MM bias and the changes due to different theoretical approaches or multipole model treatments.

The case of bis(pentacarbonylmanganese) deserves some further specific comments. As reported in Table 4 of [15], the S% contributions obtained from the PD, the multipole-modeled primary (MMPD) and experimental (MMED) densities agree very well among each other for both the Mn–C and the C–O bonds. Deviations are generally very small (about 0–2 points) and never exceed the 4 points. The Mn–Mn interaction exhibits instead a quite different behavior. The SF contribution from the Mn atom to the Mn–Mn bcp density is negative for all three densities, but the S% value for the investigated [15] PD density, −20.5%, differs significantly from that obtained after the MM refinement (−42.7%) or from the MMED (−47%). Since a large discrepancy with the PD S% estimate persists whether the MM refinement is applied to the PD or to the experimental structure factors, Lo Presti and Gatti [15] argued that its very origin is to be ascribed just to the bias introduced by the MM. By inspecting the differences in the $\nabla^2\rho$ and LS profiles along the Mn–Mn bond path, with the rp being held fixed at the Mn–Mn bcp, they could conclude that the larger S% negative contribution for the MM refined densities is basically due to a noteworthy MM bias in the $\nabla^2\rho$ distribution of the outermost core regions of the Mn basin. In fact, in the interval of the atomic L and M (s,p and not d) Mn shells, $\nabla^2\rho$ (MMPD or MMED) was found to be much less negative than $\nabla^2\rho$ (PD), which results in larger negative contributions to the bcp density from either MMPD or MMED in this interval. It was also found that refinement of the k, k' screening parameters slightly reduced the S%(PD-MMPD) difference, relative to using standard unitary parameters. However, such a difference can be much more significantly lowered, and even down to few percentage points, by diminishing the higher order of the MM expansion from $l_{max} = 4$ to $l_{max} = 1$ or even $l_{max} = 0$ (Fig. 4 of [15]). Apparently, the higher are the poles refined on the TM atoms, the larger is the MM bias for the Mn–Mn interaction. Since higher poles are, however, deemed necessary for a proper description of p and d electrons in such metals, this clearly unpleasant result could simply reveal a different problem. Diminishing of such specific MM bias by lowering the order of the MM expansion is likely the result of compensatory errors, rather than a clear signal that such an order need to be decreased, if a trustable electron density is searched for. Lo Presti and Gatti [15] also noted that in $Mn_2(CO)_{10}$ the MMPD ρ_b values are all similarly affected by the MM bias when compared to their corresponding PD reference estimates. Indeed, $|\Delta\rho_b$(MMPD-PD)| magnitudes for the Mn–Mn bcp (0.046 eÅ^{-3}) do not significantly differ from the corresponding averaged values for the Mn–C (0.069 eÅ^{-3}) and the C–O (0.025 eÅ^{-3}) bcps. However, since the electron density at bcp is about 4 and 15 times smaller at Mn–Mn than it is at the Mn–C and C–O bcps, respectively, the mentioned similarity among the MM biases on the ρ_b values does not hold true when applied on a relative scale. The corresponding percentage $\Delta\rho_b$ differences are thus very large for the Mn–Mn bond (24%) and smaller or significantly smaller for Mn–C (9%) and C–O (1%). The need to reconstruct a much lower ρ_b value is the most likely reason behind the unexpected sensitivity of the S%(Mn) value to the MM bias. Looking at the S% contributions, rather than at the SF values, emphasizes the MM bias in the case of the Mn–Mn bond, rather than dampening it as found for all other bonds in C_4F_6, CO, and

$Mn_2(CO)_{10}$ itself. One expects such a different behavior to be the rule, rather than the exception, for bcps characterized by very low electron density values. That is to say, the lower is the density at the bcp and the higher is the delocalization of sources, the lower is expected to be the otherwise impressive robustness generally observed for the S% descriptor.

5 Double Integrating the Local Source: An Unambiguous Position-Space "Population Analysis"

In Sect. 2.2, a population analysis based on the SF approach was introduced. Preliminary results for the second-row H–*X* diatomics and numerical difficulties met with this position-space electron population scheme are concisely summarized in Sect. 5.1 below. A standard SF analysis for the H–*X* series was presented earlier in Sect. 3.2.3, where the relationship between the shape of the Laplacian distributions along the series and the values of the atomic source contributions at bcp densities was highlighted. By double integrating the LS functions defined through these same Laplacian distributions, an electron population analysis matrix M, having 2×2 dimensions for diatomics, is obtained. In general, each element of M is evaluated by letting the $\mathbf{r}$ coordinate to span one atomic basin and the rp coordinate $\mathbf{r}'$ to span either this same basin (diagonal elements of M) or a different atomic basin (out-of-diagonal elements of M). Operating this way on H–*X* diatomics, one obtains the elements M(A,A) (A = H or *X*) representing the self-contributions to the atomic electron population of A, N(A), and the elements M(A,B) (A,B = H or *X* with A ≠ B) defining the contributions to the electron population of H from *X* or vice versa (13). In practice, the evaluation of the full M matrix is simpler than that. It just requires a standard 3D integration over each atomic basin as usually performed to get the various N(Ω) values, provided one stores in advance the whole set of integration grid points for all atoms in the molecule. By integrating the LS($\mathbf{r}$, $\mathbf{r}'$) within Ω with $\mathbf{r}$ running over the grid points of Ω and $\mathbf{r}'$ running over the whole set of integration grid points of all atoms, one immediately gets and separately stores the elements of a full row of the M matrix, $M(\Omega,\Omega')$; $\Omega' = 1, N$ with N being the total number of atoms in the molecule). Summing up the elements over the row equals N(Ω); hence with just one standard 3D atomic integration, one obtains both the atomic population of Ω and its decomposition in a self-contribution and in external contribution terms to this population from the remaining atoms in the system. However, this apparently simple procedure becomes often very challenging from a numerical point of view. A serious problem arises as for the "correct" assignment of atomic sources for those elements of volume integration which are characterized by very low average electron density values and electron populations (see *infra*).

5.1 Preliminary Results for the Second-Row Hydride Series and Comparisons with Mulliken's Population Analysis Scheme

The SF population analysis results for the second-row diatomics H–*X* series are detailed in Table 11, whereas Fig. 14 concisely illustrates the trends of values through the series for the elements of the population matrix *M*. A comparison with values for the corresponding terms in Mulliken's scheme [17] is also shown in the same figure. Indeed, the number of electrons N_e in the system may be written as [17, 124]:

$$N_e = \sum_{\mu\nu}^{N_{basis}} \mathbf{P}_{\mu\nu}\mathbf{S}_{\mu\nu} = \sum_{\mu\nu}^{N_{basis}} \mathbf{P}_{\mu\nu}\mathbf{S}_{\nu\mu} = tr\,(\mathbf{PS}) \tag{15}$$

Table 11 Population analysis from the double atomic basin integration of the local source function in the second-row diatomic hydrides[a]

H–*X*	$N(\Omega)$	$M(\Omega, \Omega)$[b]	$M(\Omega, \Omega')$[b]	ICP[b]	$M(\Omega, \Omega)/N(\Omega)$	IP(Ω)[c]	DEV(Ω)[c]	DEVM(Ω)[c]
H	1.901	0.23	1.65	0.39	0.12	0.081	0.063	0.063
Li	2.090	1.92	0.16		0.92	0.006	−0.009	0.036
H	1.874	0.43	1.42	0.98	0.23	0.035	0.005	0.007
Be	3.125	2.50	0.55		0.80	0.053	−0.021	0.033
H	1.787	0.67	1.09	1.48	0.38	0.024	0.000	0.004
B	4.213	3.34	0.81		0.79	0.037	−0.027	0.034
H	1.053	0.77	0.27	1.76	0.73	0.016	0.002	0.003
C	5.947	4.93	0.99		0.83	0.033	−0.002	0.031
H	0.663	0.62	0.04	1.47	0.94	0.010	0.002	0.002
N	7.337	6.46	0.86		0.88	0.028	0.010	0.030
H	0.386	0.42	−0.04	1.04	1.09	0.006	0.001	0.001
O	8.614	7.97	0.62		0.93	0.023	0.001	0.037
H	0.242	0.29	−0.05	0.71	1.20	0.004	0.000	0.001
F	9.758	9.30	0.42		0.95	0.017	−0.021	0.047

[a]HF/6-31G** wavefunction

[b]$M(\Omega, \Omega)$ represents the contribution of Ω to determining the electron population of its own basin, while $M(\Omega, \Omega')$ is the contribution to this same population from basin Ω′, with Ω′ being *X* when Ω = H and vice versa; ICP is the interchanged electron population between Ω and Ω′, ICP = $M(\Omega,\Omega') + M(\Omega', \Omega)$, that is the sum of the contribution to the electron population of Ω from Ω′ and of the contribution to the electron population of Ω′ from Ω

[c]All these parameters are related to the numerical accuracy of the population analysis. IP(Ω) represents the ignored atomic electron population, that is the atomic sum of the electron population of those volume elements within Ω that due to their too low average electron density could not be used in the evaluation of the elements of the matrix *M* (see text). The quantity DEV(Ω) is given by the atomic integration of the local deviation of the ρ reconstructed through the SF from the exact ρ value; DEVM(Ω) is given by this same atomic integration but taking the module of the integrand (see text)

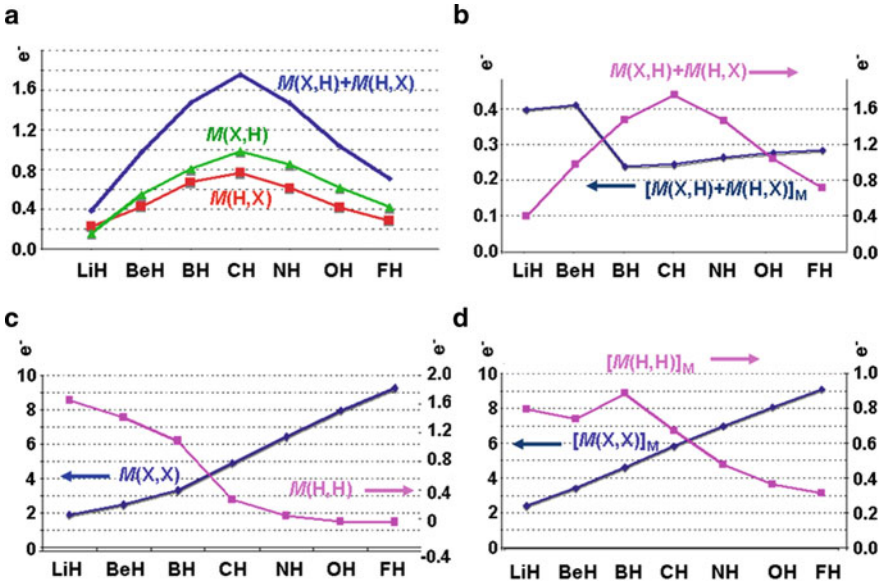


Fig. 14 Population analysis from the double atomic basin integration of the local source function in the second-row diatomic hydrides H–*X* and comparison with corresponding terms in the Mulliken's scheme. (**a**) Out-of-diagonal terms of the electron population matrix *M* as obtained from the SF approach: M (H,*X*) + M(*X*,H) = ICP, the interchanged electron population between H and *X*, given by the sum of the contribution to the electron population of H from *X* and of the contribution to the electron population of *X* from H; (**b**) comparison between the ICP as obtained from the SF approach and the corresponding overlap term $[M$ (H,*X*) + M(*X*,H)$]_{\mathrm{M}}$ in Mulliken's analysis. Note that by definition M(H,*X*) = M(*X*,H) in Mulliken's analysis; (**c**) contributions of H and *X* basins to determining their own electron populations in the SF approach; (**d**) as in (**c**) but for formally corresponding terms in the Mulliken's scheme (this figure is adapted, with permission, from material presented by the authors in [16, 33])

where N_{basis} is the number of basis set functions and **P** and **S** are the first-order density and the overlap matrix representations over that basis, respectively. These matrices are both symmetric, so that (15) may be also written as:

$$\begin{aligned} \mathrm{N_e} &= \sum_A \left[\sum_{\mu \in A, \nu \in A} \mathbf{P}_{\mu\nu}\mathbf{S}_{\mu\nu} + 2 \cdot \sum_{\mu \in A, \nu \in B \neq A} \mathbf{P}_{\mu\nu}\mathbf{S}_{\mu\nu} \right] \\ &= \sum_A \left\{ M(\mathrm{A},\mathrm{A})_{\mathrm{M}} + \sum\nolimits_{B \neq A} \left[(M(\mathrm{A},\mathrm{B}) + M(\mathrm{B},\mathrm{A})]_{\mathrm{M}} \right\}, \right. \end{aligned} \tag{16}$$

where the contribution to the number of electrons due to basis functions only centered on A is named $[M(\mathrm{A},\mathrm{A})]_{\mathrm{M}}$, and that due to one basis function centered on A and the other on B is called $[M(\mathrm{A},\mathrm{B})+M(\mathrm{B},\mathrm{A})]_{\mathrm{M}}$. The subscript "M" denotes that these quantities are defined within a Mulliken's scheme and a basis set approach. Note that $[M(\mathrm{A},\mathrm{B})+M(\mathrm{B},\mathrm{A})]_{\mathrm{M}}$ corresponds to the Mulliken's total

overlap population between atoms A and B [124]. Due to their definition, it is natural to make a formal association between these quantities and the corresponding elements of the matrix M defined through the SF approach.

Figure 14a shows that the interchanged population (ICP) $= M(X,\mathrm{H}) + M(\mathrm{H},X)$ increases with increasing covalency along the series. It so reaches a maximum of about 1.7 e^- at $X = \mathrm{C}$ and then monotonically decreases to about 0.8 e^- at $X = \mathrm{F}$. Note that the curve is not symmetric with respect to $X = \mathrm{C}$, the central member of the series, because, at variance with Li–H, the H–F molecule retains a significant covalent character (compare the Laplacian distributions in Fig. 3).

The M matrix defined through the SF approach is not symmetric and therefore the two curves $M(X,\mathrm{H})$ and $M(\mathrm{H},X)$ differ between each other with the contribution from X to the population of H exceeding that of H to X, except for LiH. It is pleasing that the external contributions to the population follow a similar trend, both increasing/decreasing with covalency increase/decrease through the series. Inspection of Fig. 3 easily explains why the two external contributions need to be generally different. Using their expression given by the surface term in (12), one observes that both contributions are evaluated on the same X| H interatomic surface and using the same electron density on this surface. However, the two contributions do not simply and necessarily differ in sign as would naively suggest the opposite direction of their normals to the surface, but differ in value because the electron density on the surface is weighted by a term involving the gradient of the scalar $1/|\mathbf{r} - \mathbf{r}_s|$. This term reflects both the size and shape of the basin being integrated which are in turn related also to the convex or concave nature of the interatomic surface (IAS). Concave shapes are typical of anionic-like atomic basins, whereas convex ones characterize those of cations. The contribution of one atom to the population of an atom bonded to it and the magnitude of such contribution relative to the corresponding one from the linked atom is therefore deeply related to the nature of the associated bonding interaction. Contrary to what observed for other members of the series, the H-basin in LiH is largely anionic, exhibits a clearly concave IAS, and thus found to contribute more to the Li population than Li does to that of the H atom.

Figure 14b compares the trend of the ICP values with those of the formally associated overlap Mulliken's populations $[M(X,\mathrm{H}) + M(\mathrm{H},X)]_\mathrm{M}$. It is gratifying that the former values appear quite more realistic. They exhibit much larger and chemically reasonable values in general, and with comparatively larger values for the shared interactions than for the ionic ones. The Mulliken's overlap populations, instead, show not only a quite erratic behavior but also an unrealistic larger overlap for ionic than for shared interactions. Figure 14c, d compares the self-population contributions in the two population schemes. The $M(\mathrm{H},\mathrm{H})$ values decrease with increasing X electronegativity and show a clearly sharper decrease in the central part of the graph where the nature of the H changes from anionic to cationic, whereas the $M(\mathrm{H},\mathrm{H})_\mathrm{M}$ values exhibit an unexpected peak for $X = \mathrm{B}$ before decreasing with increasing electronegativity of X. The $M(X,X)_\mathrm{M}$ values, instead, show the expected monotonic increase with increasing Z, analogously to the $M(X,X)$ values.

The preliminary results obtained using this position-space unambiguous population analysis look indeed promising. Apart from a conference talk [16] and a MSc thesis [33], they are presented here for the first time in the literature. Several other systems have been tested by us, including a number of polyatomic molecules, but, despite its undoubtedly physical appeal, a general application of the approach appears at the moment unfeasible. The method is in fact often numerically instable, besides being very time consuming. While one always recovers with great accuracy (13) the atomic electron population $N(\Omega)$ in terms of the sum of the internal, $N_i(\Omega)$, and external contributions, $N_o(\Omega)$, evaluating the relative weight of such contributions may often become a true nightmare because of the difficulties inherent to the SF reconstruction of the electron density in regions characterized by extremely low density values. As mentioned in Sect. 2.3, when $\rho(\mathbf{r})$ falls below 10^{-4} au the density reconstruction through (2) is problematic. If such a difficulty simply limits the choice of possible rps in standard SF analysis, it makes instead extremely challenging the precise atomic assignment of sources for the electron population of the volume elements characterized by very low density. These elements may play a dominant role since they generally cover a very large portion of the atomic basin, and their cumulative electron population is often far from being negligible. We have explored many viable strategies to afford this problem. A first, apparently obvious solution is to not assign any source when the electron density at a given grid integration point turns out to be poorly reconstructed [large *ER%* value in (14)]. However, due to the unbounded nature of the atomic sources – they may be either positive or negative – one often encounters the case of a very good reconstruction of the density (very small *ER%* values) in spite of unacceptable uncertainties in the values of the composing sources. As a consequence, the relative weight of the internal and external contributions may become quite unstable and disappointingly sensitive to the choice of otherwise irrelevant computational parameters. The most successful approach devised thus far is to set a threshold on the lowest accepted density value for a safe reconstruction of the density and to not assign the electron population of those volume elements associated with the grid integration points having a density value below such a threshold. Although the resulting total ignored population (IP) was never found to be negligible, it was yet thought to be "acceptable" in several cases. Using a threshold on ρ of 5×10^{-4} au, $IP(\Omega)$ values for the H–*X* series never exceeds 0.08 e^- and are in most cases as low as or even lower than 0.01 e^- (see Table 11). Usually, the threshold on ρ was chosen as the lowest possible ρ value for which the relative weight of internal and external contributions to $N(\Omega)$ remains stable. The overall quality of the atomic electron population decomposition (11) was then judged by integrating over Ω the local deviation of the reconstructed density from the exact ρ value, for those integration points with ρ values above the selected threshold. The value of such an integral and of that obtained by taking the module of the integrand are respectively reported as $DEV(\Omega)$ and $DEVM(\Omega)$ in Table 11. Both quantities were generally tolerably small in magnitude for the H–*X* series, the larger values not exceeding 0.06 e^-.

On the whole, the population analysis based on the SF appears promising, but it is to be considered in its infancy as for a general application. At present, we are unable to say whether it is just conceptually relevant or also practically feasible. As discussed in Sect. 2.2, it would represent a unique, unambiguous tool for comparing populations derived from experimental and theoretical densities.

6 Orbital and Core–Valence Decompositions of the Atomic SF Contributions

In a very recent paper [12], entitled *On the Interpretation of the Source Function*, Farrugia and Macchi critically analyzed the "chemical" information present in the SF in several case studies. In the following, we concisely illustrate their adopted approach and review some of the results obtained along with the main conclusions which were drawn from such study. We also discuss a number of issues that in our opinion would deserve a different interpretation and propose alternative, possibly more apt ways for partitioning the atomic SF.

Farrugia and Macchi (F&M) were essentially aimed at verifying "whether the SF in fact carries information comparable with other well-established decomposition schemes that relate, more or less straightforwardly, to commonly accepted chemical concepts," in particular the electron delocalization. A quantitative and physically sound measure of the latter is provided by the delocalization indices, which yields the average fractional number of electron pairs shared between two atomic basins, as discussed repeatedly in this review. As no formal relationship exists between the SF and the delocalization indices [14], F&M proposed to test it inductively by comparing these two indicators in terms of corresponding composing contributions. Namely, in the case of electron densities derived from DFT calculations, they examined how the SF and the $\delta(\Omega,\Omega')$ decompose in terms of the canonical valence and core Kohn-Sham MOs, or also in terms of the individual MOs.[32] For the "synthetic" or experimental densities, given in terms of complex static X-ray structure factors, delocalization indices are generally unavailable, unless using the wavefunction constrained method [100] or, in principle, one of the proposed forms of density matrix refinements ([126] and references therein). A core/valence decomposition of the SF, formally analogous to that performed on DFT densities, is anyhow possible. F&M used the XD2006 program suite [59] and the Hansen–Coppens multipole formalism [10] to obtain and decompose the aspherical electron density of each atom in terms of a spherical core and of an aspherical (spherical + deformation) valence density, from which the corresponding contributions to the atomic SF were calculated.

[32]A similar SF decomposition, in terms of σ and π MOs, was reported earlier in this chapter, (Sect. 3.2.2), when discussing whether the SF may in some way reflect π-electron conjugation.

As a first example, F&M examined, at the theoretical level, the case of the C_{3v} borane carbonyl BH_3CO, a simple model for carbonyl coordination to a strong Lewis acid and, by extension, to transition metals. A comparison was made first between the MO's decomposition of the B–C, C–O, and B–H bcp densities, and that of the delocalization indices for these same pairs of atoms. By definition, the electron density at a point **r**, $\rho(\mathbf{r})$, equals S(**r,** all space), the SF contribution from the whole space, with the rp taken at **r**. While it was noted that the two descriptors δ(B,H) and S($bcp_{B–H}$, all space) have very similar overall MO contributions, this was obviously not the case for the two bonds (C–O and B–C) whose bcp lies on the symmetry axis and which have π-components. For instance, the major contributions to δ(C,O) come from an E pair of π-bonding MOs essentially localized around the C–O bond, whereas these same orbitals, because of their symmetry, cannot clearly contribute to ρ_b(C–O), and hence to S($bcp_{C–O}$, all space). When the MO's decomposition is applied to the SF contributions from the atomic basins, a very similar picture is obtained, since the E pair of π-bonding MOs localized around the C–O bond yields again a negligible, though nonzero contribution, to ρ_b(C–O).[33] F&M thus argue that, at variance with δ(C,O), essentially "no information about the extent of π-bonding or π-back donation is contained in the SF at $\mathbf{r}_b$(C–O)" [12]. This observation on C–O was then generalized by observing that "when the rp is close to the nodal plane of an orbital, this orbital makes a low to negligible contribution to the SF which has clear implications for the interpretation of π-interactions" [12]. Of course, we agree with the indirect suggestion by F&M that moving the rp out from the nodal plane enables such an orbital to enhance its contribution to the SF, so providing insights on the π-bonding mechanisms, as we demonstrated for benzene in Sect. 3.2.2. However, while the F&M observation is correct in a, let to say, "zero-order" approach (direct effect), it is no longer so if higher-order, indirect effects are considered. Indeed, "essential to the orbital theory of electronic structure is the property of self-consistency – that each orbital be determined by its interaction with the average Coulomb and exchange potential generated by electrons in the other occupied orbitals. Thus the density distributions derived from σ and π orbitals are not independent of one another" ([6]; p. 76) at any point of the molecular space. The interdependence of σ and π electron distributions is a well-known, documented fact in the literature [20, 127] and was already exploited earlier in this chapter (Sect. 3.2.2) when analyzing whether the SF contributions may be affected in some way by π-electron conjugation when the rp lies in the π-nodal plane. As said earlier, if a bcp lies in the nodal surface of a π-orbital, this latter will be unable to provide a *direct* contribution to the density at that point, but it will supply an *indirect* one through its interaction with the other orbitals, including those of σ-symmetry, able to contribute to the electron density at bcp. A very simple *gedanken* example on the $C_2H_4^{+n}$ (n = 0–4) series illustrates

[33] It is the integration over the whole space that results in a null contribution to the ρ_b(C–O) density from the π-bonding MOs. Separate integration over the atomic basins may yield non-zero values (typically less than 1% of the $|\rho_b|$ value) which are constrained to sum up to zero.

this behavior. Although the π-orbitals do not contribute to the electron density at the C–C bcp, removing electrons from them while keeping the geometry fixed and allowing for self-consistency has a great effect on the contributions from the σ-orbitals. To compensate for the increasing positive charge and the concomitant lack of occupied bonding π-orbitals along the series, the σ-ones are found to increase their overall contribution at the bcp, hence to ρ_b, despite the total number of electron in the system is being drastically decreased through the series. The increase is relevant, ρ_b being 0.344 au for $n = 0$, rising to 0.361 and 0.381 au for $n = 1$ and 2 and reaching 0.395 au for $n = 4$ (DFT/B3LYP 6-311++G** level). All in all, the contribution of π-orbitals and π-bonding at points lying on the nodal surface may be revealed to some extent by the SF through the effect the π-orbitals have on the values of the σ-orbitals contributions at these points. The SF description for rps lying on the nodal surface will be clearly unchanged whether an electron density given in terms of MOs or an equivalent one, like for instance a numerical electron density obtained from maximum entropy method for which σ/π separation is unavailable, is used. Both densities will contain those physical effects that are translated into the σ–π model using a MO formulation, and the SF will simply reflect such effects, regardless a σ/π decomposition may be actually realized or not. The merit of the MO model and of the SF decomposition in terms of MOs is that these approaches solicit and drive to explore rps other than the bcps and out from the nodal plane(s), and then they offer a rationale for interpreting the resulting, often significantly varied, atomic SF patterns (see, e.g., Sect. 3.2.2). Hence, although it is in principle false that "the SF taken at the C–O bcp cannot provide any information about the extent of C–O π-bonding" [12], one may anticipate a large enhancement of the effect of such bonding scheme on the atomic SF values when the rp is moved out from the π-orbital nodal plane.

F&M also noted that, at variance from the localization indices, a substantial contribution to the SF at the C–O bcp comes from the core electrons. Percentage orbital contributions to δ(C, O) from the three MO core orbitals is less than 1%, whereas the MO core orbital 2, which is essentially the unhybrized 1*s* orbital of C, yields the dominant source from C (19.0% out of a total of 39.8%) and a non-negligible negative sink from O (−12.5% out of a total of 58.4%). According to F&M, this "reinforces the idea that the $\delta(\Omega, \Omega')$ is more closely reproducing the concepts of electron sharing" [12]. While we clearly have no doubt that $\delta(\Omega, \Omega')$ are more intimately connected to electron sharing, we also believe that one should never ask to a descriptor under exam, i.e., the SF in this case, to comply with what one would expect to deduct from it, but rather focus on what it is actually observed. As we will discuss below, the large source and sink arising from the core orbital 2 when this orbital is respectively integrated over the C and O atoms is a natural, physical consequence of the large electronegativity difference of these two atoms and of the consequent large shift of the C–O bcp position toward the more electropositive C. The net charges of C and O in BH_3CO amount to +0.88 and −1.09 e^-, respectively. The bcp lies so close to the *atomic* core-shell depletion region of C that the C 1*s* core MO contribution to the bcp density from C will be totally different from that found for standard, nonpolar C–C bonds in hydrocarbons.

Indeed, F&M found that in acetamide the C–C bcp density is clearly dominated by the valence density, with core contributions lower than 1%. The bcp in C–O lies at 0.39 Å from the C nucleus and, as mentioned above, within the *atomic K* shell (core) depletion region of C which ends at about 0.43 Å from the C nucleus, whereas for a less polarized bond the bcp is normally located at a much larger distance from the C nucleus (0.6–0.75 Å in hydrocarbons) and well inside the region of the *atomic* VSCC of carbon. The SF from C and the large magnitude of the core "contribution" to the SF from C takes precisely into account this difference. When integrated over the O basin, the core orbital 2 yields instead a negative source at bcp because this basin may only include that outer part of the C 1*s* MO where the associated Laplacian is positive.[34] The "odd" contributions from the C 1s MO should therefore not be taken as a "problem" of the SF, but rather as a very interesting property of this approach. The SF demonstrates to be capable of discriminating the way a given atom determines the density at its various bcps as a function of the relative electronegativity of the atoms to which it is bonded.

In particular, one should never forget that the SF yields two, usually different, values for the density contribution of two bonded atoms Ω–Ω′ at their intervening bcp, while the number of shared pair of electrons between these atoms is represented by just a single $\delta(\Omega,\Omega')$ value. On top of not being physically related among them in a direct way, delocalization indices and SF values need to differ in given circumstances. For instance, the latter are by nature able to distinguish whether two bonded atoms are playing similar or quite distinct roles in their bonding interaction.

F&M properly recognize that, while useful for relating the SF to chemical concepts expressed in a MO framework model, the decomposition of an observable like ρ in terms of molecular orbitals is arbitrary and subject to the particular choice of MO settings (canonical, localized, etc). As an obvious consequence, the SF decomposition in terms of MOs is arbitrary as well. Because of this, F&M propose to examine the decomposition of the atomic SF into core and valence contributions, as another, potentially more useful, approach. In the case of BH_3CO, which is characterized by largely polar bonds, the core densities provide important contributions from individual basins toward the total SF, but for other two investigated systems (acetamide and thiocumarin), including second and third period atoms and generally much less polar interactions, the valence density was found to provide the determining contributions, as for the delocalization indices. The only notable exception was for the polarized C–O and C–S bonds, which further confirms the interpretation given earlier for the core contributions in borane carbonyl. For heavier elements such as the iron atom in $Fe(CO)_5$, the core density plays a significant role, if a large [Ar] core description is used for such a density. Indeed the [Ar] core density provides a source from iron which amounts to about 40/49% of the Fe–C bcp density, whereas the valence part acts as a moderate sink,

[34]The large negative source obtained at the bcp density when the C 1*s* MO is integrated over the O basin demonstrates that such orbital cannot be longer safely identified as a C 1*s* MO in a molecular context. This is just an example of the limits of the MO model interpretation.

−8.7/−18.6% (the two entries refer to the equatorial and axial bonds, respectively). But if the small [Ne] core description of the valence is adopted (as customarily done when performing careful effective core potential calculations on these 3*d* transition elements), the core contribution becomes almost negligible, since the overall Fe basin contribution for the Fe–C bcp density derives almost entirely from the relatively diffuse 3*s* and 3*p* density. This clearly opens the problem of what should be considered the "real" core for these elements, since different core/valence partitionings of MOs lead to completely different conclusions. The fact that a small [Ne] core is suggested in several PP schemes for the 3*d* elements seems to comply with a nonnegligible role in bonding of the 3*s* and 3*p* metal electrons. F&M report that "the role of (n − 1)*s* and (*n* − 1)*p* orbitals in the electron sharing of transition metal complexes is generally quite negligible, according to many partitioning schemes" [12]. However, although referred to the metal–metal rather than to the metal–ligand interactions in these complexes, two evidences, already discussed in this chapter, seem in contrast with such a claim. The first is the relative depletion of the atomic (*n* – 1) (*M*-shell for Fe) charge concentration shell which we observed to occur when the metal–metal bond is allowed to form (Sect. 3.4.2), suggesting that metal–metal bonding in these complexes may imply a substantial rearrangement in the (*n* − 1) shell. The second evidence is the very large importance of the metal–metal core covariance, which was found using the ELF analysis on the M_2(formamidinate)$_4$ complexes and which denotes that the fluctuations of the core basin populations of the two metal are highly correlated one to another [85]. The ELF core basins include the (n – 1)*s* and (n – 1)*p* electrons, besides most of the *d* electrons. On the whole, it looks like that a precise assignment to core of those electrons that do not participate to bonding is quite challenging in the case of the TM elements, and that this choice may depend on the way bonding is analyzed (position-space descriptors vs. orbital models).

F&M conclude their paper [12] by recommending caution in associating some chemical concepts with features of the SF and especially so for heavier elements. While they admit that the "SF might contain some information about chemical bonding," they also found that "it is often distinct from that of electron delocalization between two or more atomic basins" and they accordingly warn the reader that its interpretation may not be so straightforward. We mostly agree on these points and in particular on the need of a thoughtful use of the SF. In this section, we also examined at length the issues F&M believe should be considered by scientists making use of the SF for an electron density analysis. We concur, for instance, on the need to explore other rps besides the bcps in given circumstances and on the observation that the SF contributions may emphasize the role of internal electrons because of the significant shift of the bcp toward one atom in a bonded pair typical of polar bonds.

We wish to recall, however, that any functional decomposition of the electron density is intimately related to the way it is calculated, and it is therefore essentially arbitrary. Even the interpretive useful σ/π separation is tied to the LCAO-MO formalism. When using a density on a grid, one will never be able to discriminate the σ and π contributions/effects, but will simply register what happens to the total

density (in terms of various position-space descriptors, including the density values at the bcps, the density curvatures at these same points, the SF contributions, etc.) when the chemical environment changes, in a series of related systems. One may then interpret these same density changes using an equivalent electron density which allows for a σ/π separation. But, despite its usefulness, this will be just one out of the many possible interpretations. Analogously, a similar, though more limited, arbitrariness would exist for the core/valence separation when performed on a functional space.

On the whole, when commenting the ability of the SF descriptor to "contain some information about chemical bonding," we should always be cautious that one is basing his or her judgment on a preconception of bonding which is derived from given, more or less arbitrary, models. A more correct position is perhaps that of observing and accepting what the SF descriptor is telling us in the various circumstances and of not blaming its possible inhability to comply with "commonly accepted chemical concepts" based on such models. It is also true, however, that comparison with other physically well-defined descriptors (e.g., delocalization indices) or even interpretation through arbitary formalisms (e.g., MOs) may enhance our understanding of the SF descriptor and of how to make a better use of it.

6.1 Alternative Decompositions of the Atomic SF Contributions and of Localization and Delocalization Indices

The possible decompositions of the atomic SF are clearly not only limited to those performed in the functional space in which the density is represented. Since the SF is defined on the basis of an observable in-position space, the most natural subpartitioning of the atomic SF contributions should probably also be realized in such a space. Indeed, it is well known that canonical core MOs have tails in the valence MO's regions and vice versa, which leads to significant local overlaps and may clearly affect the interpretation of the resulting decomposition in SF contributions. We are currently exploring an atomic-shell subpartitioning of the atomic SF contributions in position space, using the shell radii derived from the atomic Laplacian distributions. Clearly, such an approach represents the best way to compare experimental and theoretical decompositions of the SF atomic contributions, since any dependence from the functional space used to express the electron density is avoided. By operating this way, a core/valence decomposition of the atomic SF contributions in the position space becomes immediately available (and, in case of TM elements, for any large- or small-core description one is interested to). If the atomic-shell subpartitioning is also applied to the delocalization indices, one may compare these latter with the atomic SF contributions, and within a similar position-space decomposition scheme. Quite interestingly, when the partitioning of the localization or delocalization indices is fully carried out in the position rather

than in the combined position and functional spaces, mixed terms related to one electron being located in one shell when the other is in another shell occur. For instance, a $\delta(\Omega,\Omega')$ delocalization index can be partitioned in four contributions, two of which are mixed core–valence terms, while the other two are the core–core and the valence–valence terms. If not negligible, the mixed core–valence terms enhance the importance of the core in yielding a given $\delta(\Omega,\Omega')$ value, an effect which cannot be revealed by a standard MO decomposition of the delocalization indexes.

Recently, Francisco et al. [128] demonstrated that the one-electron functions derived from the diagonalization of the Fermi hole averaged over an atomic basin Ω have quite relevant properties. When the eigenvalues of these DAFH orbitals [92] are close to 1.0 or 0.0, these orbitals are almost fully localized in Ω [128] or in its complementary space, and can be simply ignored in computing $\delta(\Omega,\Omega')$, because they do not contribute to the chemical bonding between the two fragments. On the other hand, eigenvalues close to 0.5 correspond to maximally delocalized orbitals that participate significantly to the bonding. In a way, when properly ordered in terms of their eigenvalues, the DAFH orbitals enable one to obtain the fastest convergent expression for $\delta(\Omega,\Omega')$. A comparison of the atomic SF contributions and delocalization indices in terms of a partitioning based on DAFH orbitals could represent a particularly suited orbital-based approach to enhance our understanding of the physical difference between these two descriptors.

The alternative decomposition schemes we have put forth in this session lack any arbitrariness inherent to the partitioning based on the MOs, and they will be both explored. While the first is applicable to both experimental and theoretical densities, the second requires the knowledge of the pair density matrix in some form.

Acknowledgments I am deeply indepted to and warmly thank Richard Bader for his fundamental contribution to the seminal work on the Source Function. I also thank Luca Bertini, Fausto Cargnoni, Davide Lasi, and Leonardo Lo Presti for their precious collaboration in developing and applying the SF. I thank the Danish National Research Foundation for partial funding of this work through the Center for Materials Crystallography (CMC).

References

1. Bader RFW, Gatti C (1998) A Green's function for the density. Chem Phys Lett 287:233–238
2. Gatti C (2005) Chemical bonding in crystals: new directions. Z Kristallogr 220:399–457
3. Bertini L, Cargnoni F, Gatti C (2007) Chemical insight into electron density and wave functions: software developments and applications to crystals, molecular complexes and materials science. Theor Chem Acc 117:847–884
4. Gatti C (2007) Solid state applications of QTAIM and the source function – Molecular crystals, surfaces, host-guest systems and molecular complexes. In: Matta CF, Boyd RJ (eds) The quantum theory of atoms in molecules. From solid state to DNA and drug design, 1st edn. Wiley, Weinheim

5. Arfken G (1985) Mathematical methods for physicists. Academic, Orlando, FL
6. Bader RFW (1990) Atoms in molecules: a quantum theory. In: International series of monographs on chemistry, vol 22. Oxford Science, Oxford
7. Bader RFW (1998) A bond path: a universal indicator of bonded interactions. J Phys Chem A 102:7314–7323
8. Pendás AM, Francisco E, Blanco MA, Gatti C (2007) Bond paths as priviliged exchange channels. Chem Eur J 13:9362–9371
9. Gatti C, Cargnoni F, Bertini L (2003) Chemical information from the source function. J Comput Chem 24:422–436
10. Coppens P (1997) X-ray charge densities and chemical bonding. In: IUCr texts on crystallography, vol 4. International Union of Crystallography, Oxford University Press, Oxford
11. Morse PM, Feshbach H (1981) Methods of theoretical physics, part I, chap 1. Feshbach, Minneapolis
12. Farrugia LJ, Macchi P (2009) On the interpretation of the source function. J Phys Chem A 113:10058–10067
13. Gatti C, Bertini L (2004) The local form of the source function as a fingerprint of strong and weak intra- and intermolecular interactions. Acta Crystallogr A60:438–449
14. Gatti C, Lasi D (2007) Source function description of metal–metal bonding in d-block organometallic compounds. Faraday Discuss 135:55–78
15. Lo Presti L, Gatti C (2009) Using the source function descriptor to dampen the multipole bias in charge density studies from X-ray structure factors refinements. Chem Phys Lett 476:308–316
16. Lasi D, Gatti C (2006) Unambigous electron population analysis from the electron density observable? ECDM-IV, p 25 abstract, Brandenburg on the Havel, Germany, 26–29 January 2006
17. Mulliken RS (1955) Electronic population analysis on LCAO-MO molecular wave functions I. J Chem Phys 23:1833–1846
18. Spackman MA, Byrom PG (1997) A novel definition of a molecule in a crystal. Chem Phys Lett 267:215–220
19. Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. Theor Chim Acta 44:129–138
20. Gatti C, Ponti A et al (1992) Nitranions and their precursors: charge density rearrangements and 15N NMR chemical shift changes. J Am Chem Soc 114:8634–8644
21. Discussion G (2007) Chemical concepts from quantum mechanics. Faraday Discuss 135:125–149
22. Bultinck P (2007) Critical analysis of the local aromaticity concept in polyaromatic hydrocarbons. Faraday Discuss 135:347–365
23. Gatti C, Fantucci P, Pacchioni G (1987) Charge density topological study of bonding in lithium clusters. Theor Chim Acta 72:433–458
24. Cao WL, Gatti C, MacDougall PJ, Bader RFW (1987) On the presence of non-nuclear attractors in the charge distributions of Li and Na clusters. Chem Phys Lett 141:380–385
25. Bader RFW, Bayles D (2000) Properties of atoms in molecules: group additivity. J Phys Chem A 104:5579–5589
26. Bader RFW (2009) Nearsightedness of electronic matter as seen by a physicist and a chemist. J Phys Chem A 112:13717–13728
27. Matta CF, Bader RFW (2003) Atoms-in-molecules study of the genetically-encoded amino acids. III. Bond and atomic properties and their correlations with experiment including mutation-induced changes in protein stability and genetic coding. Proteins Struct Funct Genet 52:360–399
28. Bader RFW, Beddall PM (1972) Virial field relationships for molecular charge distributions and the spatial partitioning of molecular properties. Chem Phys 56:3320–3329
29. Srebrenik S, Bader RFW (1975) Towards the development of the quantum mechanics of a subspace. J Chem Phys 63:3945–3961

30. Benson SW, Cruichshank FR et al (1969) Additivity rules for the estimation of thermochemical properties. Chem Rev 69:279–324
31. Mladenovic M, Arnone M et al (2009) Environmental effects on charge densities of biologically active molecules: do molecule crystal environments indeed approximate protein surroundings? J Phys Chem B 113:5072–5082
32. Engels B, Schmidt TC et al (2010) Challenging problems in charge density determination. Struct Bond, in press
33. Lasi D (2006) MSc thesis, Milano University, Milano
34. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 38:3098–3100
35. Perdew JP (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys Rev B 33:8822–8824
36. Perdew JP (1986) Erratum: Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys Rev B 34:7406
37. Dunning TH, Hay PJ (1976) In: Schaefer III HF (ed) Modern theoretical chemistry, vol. 3. Plenum, New York, p 1
38. Fradera X, Austen MA, Bader RFW (1999) The Lewis model and beyond. J Phys Chem A 103:304–314
39. Hô M, Smith VH et al (1998) Molecular similarity based on information entropies and distances. J Chem Phys 108:5469–5475
40. Jeffrey GA (1997) An introduction to hydrogen bonding. Oxford University Press, New York
41. Carroll MT, Bader RFW (1988) An analysis of the hydrogen bond in BASE-HF complexes using the theory of atoms in molecules. Mol Phys 65:695–722
42. Espinosa E, Molins E, Lecomte C (1998) Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. Chem Phys Lett 285:170–173
43. Koch U, Popelier PLA (1995) Characterization of CHO hydrogen bonds on the basis of the charge density. J Phys Chem 99:9747–9754
44. Gatti C, May E et al (2002) Fundamental properties and nature of CH··O interactions in crystals on the basis of experimental and theoretical charge densities. The case of 3, 4-Bis (dimethylamino)-3-cyclobutene-1, 2-dione (DMACB) crystal. J Phys Chem A 106: 2707–2720
45. Fuster F, Silvi B (2000) Does the topological approach characterize the hydrogen bond? Theor Chem Acc 104:13–21
46. Martín Pendás A, Blanco MA, Francisco E (2006) The nature of the hydrogen bond: a synthesis from the interacting quantum atoms picture. J Chem Phys 125:184112
47. Overgaard J, Schiøtt B et al (2001) The charge density distribution in a model compound of the catalytic triad in serine proteases. Chem Eur J 7:3756–3767
48. Gilli G, Gilli P (2000) Towards an unified hydrogen-bond theory. J Mol Struct 552:1–15
49. Sørensen J, Clausen HF et al (2007) Short strong hydrogen bonds in 2-acetyl-1, 8-dihydroxy-3, 6-dimethylnaphthalene: an outlier to current hydrogen bonding theory? J Phys Chem A 111:345–351
50. Schmidtmann M, Farrugia LJ et al (2009) Experimental and theoretical charge density study of polymorphic isonicotinamide-oxalic acid molecular complexes with strong O· · ·H· · ·N hydrogen bonds. J Phys Chem A 113:13985–13997
51. Jeffrey GA, Saenger W (1991) Hydrogen bonding in biological structures. Springer, Berlin
52. Gilli G, Bellucci F et al (1989) Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the β-diketone fragment. J Am Chem Soc 111:1023–1028
53. Bertolasi V, Gilli P et al (1991) Evidence for resonance-assisted hydrogen bonding. 2. Intercorrelation between crystal structure and spectroscopic parameters in eight intramolecularly hydrogen bonded 1, 3-diaryl-1, 3-propanedione enols. J Am Chem Soc 113: 4917–4925

54. Becke AD, Edgecombe KE (1990) A simple measure of electron localization in atomic and molecular systems. J Chem Phys 92:5397–5403
55. Savin A, Nesper R et al (1997) ELF: the electron localization function. Angew Chem Int Ed Engl 36:1809–1832
56. Savin A, Silvi B (1994) Classification of chemical bonds based on topological analysis of electron localization functions. Nature 371:683–686
57. Bertolasi V, Pretto L et al (2006) π-Bond cooperativity and anticooperativity effects in resonance-assisted hydrogen bonds (RAHBs). Acta Crystallogr B62:850–863
58. Madsen GKH (1999) Ph.D. thesis, Aarhus University, Denmark
59. Volkov A, Macchi P et al (2006) XD2006 – a computer program package for multipole refinement, topological analysis of charge densities and evaluation of intermolecular energies from experimental and theoretical structure factors. Buffalo University, USA. http://xd.chem.buffalo.edu/
60. Ponec R, Gatti C (2009) Do the structural changes defined by the electron density topology necessarily affect the picture of the bonding? Inorg Chem 48:11024–11031
61. Gatti C, Saunders VR, Roetti C (1994) Crystal field effects on the topological properties of the electron density in molecular crystals: the case of urea. J Chem Phys 101:10686–10696
62. May E, Destro R, Gatti C (2001) The unexpected and large enhancement of the dipole moment in the 3, 4-Bis(dimethylamino)-3-cyclobutene-1, 2-dione (DMACB) molecule upon crystallization: a new role of the intermolecular CH–O interactions. J Am Chem Soc 123:12248–12254
63. Krijn MPCM, Graafsma H, Feil D (1988) The influence of intermolecular interactions on the electron-density distribution. A comparison of experimental and theoretical results for α-oxalic acid dehydrate. Acta Crystallogr B 44:609–616
64. Spackman MA, Byron PG et al (1999) Influence of intermolecular interactions on multipole-refined electron densities. Acta Crystallogr A 55:30–47
65. Ponec R, Lendvay G, Chaves J (2008) Structure and bonding in binuclear metal carbonyls from the analysis of domain averaged Fermi Holes. I. $Fe_2(CO)_9$ and $Co_2(CO)_8$. J Comput Chem 29:1387–1398
66. Macchi P, Sironi A (2003) Chemical bonding in transition metal carbonyl clusters: complementary analysis of theoretical and experimental electron densities. Coord Chem Rev 238–239:383–412
67. Gervasio G, Bianchi R, Marabello D (2004) About the topological classification of the metal-metal bond. Chem Phys Lett 387:481–484
68. Macchi P, Garlaschelli L, Sironi A (2002) Electron density of semi-bridging carbonyls. metamorphosis of CO ligands observed via experimental and theoretical investigations on $[FeCo(CO)_8]^-$. J Am Chem Soc 124:14173–14184
69. Macchi P (2009) Electron density distribution in organometallic materials. Chimia 63:1–6
70. Bianchi R, Gervasio G, Marabello D (2000) Experimental electron density analysis of $Mn_2(CO)_{10}$: metal−metal and metal−ligand bond characterization. Inorg Chem 39: 2360–2366
71. Farrugia LJ, Mallinson PR, Stewart B (2003) Experimental charge density in the transition metal complex $Mn_2(CO)_{10}$: a comparative study. Acta Crystallogr B 59:234–247
72. Farrugia LJ, Cameron E, Tegel M (2006) Chemical bonds without "chemical bonding"? A combined experimental and theoretical charge density study on an iron trimethylene-methane complex. J Phys Chem A 110:7952–7961
73. Scherer W, McGrady GS (2004) Agostic interactions in d(0) metal alkyl complexes. Angew Chem Int Ed 43:1782–1806
74. Scherer W, Sirsch P et al (2003) Valence charge concentrations, electron delocalization and β-agostic bonding in d(0) metal alkyl complexes. Chem Eur J 9:6057–6070
75. Reisinger A, Trapp N et al (2007) Homoleptic Silver(I) acetylene complexes. Angew Chem Int Ed 46:8295–8298

76. Scherer W, Wolstenholme DJ et al (2010) On the nature of agostic interactions in transition-metal amido complexes. Angew Chem Int Ed 49:2242–2246
77. Flierler U, Burzler M et al (2008) Electron-density investigation of metal-metal bonding in the dinuclear "Borylene" complex [{$Cp(CO)_2Mn$}$_2$(μ-B*t*Bu)]. Angew Chem Int Ed 47:4321–4325
78. Goetz K, Kaupp M et al (2009) Comparative analysis of electron-density and electron-localization function for dinuclear manganese complexes with bridging boron- and carbon-centered ligands. Chem Eur J 15:623–632
79. Poulsen RD, Jorgensen MRV et al (2007) Synchrotron X-ray charge-density study of coordination polymer $[Mn(HCOO)_2(H_2O)_2]^{\infty}$. Chem Eur J 13:9775–9790
80. Overgaard J, Clausen HF et al (2008) Experimental and theoretical charge density study of chemical bonding in a Co dimer complex. J Am Chem Soc 130:3834–3843
81. Farrugia LJ, Middlemiss DS et al (2008) A combined experimental and theoretical charge density study of the chemical bonding and magnetism in 3-amino-propanolato Cu(II) complexes containing weakly coordinated anions. J Phys Chem A 112:9050–9067
82. Overgaard J, Cameron J et al (2009) Experimental electron density study of the Mg-Mg bonding character in a magnesium(I) dimer. J Am Chem Soc 131:4208–4209
83. Farrugia LJ, Cameron E et al (2009) The QTAIM approach to chemical bonding between transition metals and carbocyclic rings: a combined experimental and theoretical study of (η^5-C_5H_5)Mn(CO)$_3$, (η^6-C_6H_6)Cr(CO)$_3$, and (*E*)-{(η^5-C_5H_4)CF-CF(η^5-C^5H^4)}(η^5-C^5H^5)$_2$Fe$_2$. J Am Chem Soc 131:1251–1268
84. McGrady GS, Sirsch P et al (2009) Nature of the bonding in metal-silane σ-complexes. Inorg Chem 48:1588–1598
85. Llusar R, Beltrán A et al (2001) Topological analysis of multiple metal−metal bonds in dimers of the M_2(Formamidinate)$_4$ type with M = Nb, Mo, Tc, Ru, Rh, and Pd. J Phys Chem A 105:9460–9466
86. Ponec R, Yuzhakov G, Sundberg MR (2005) Chemical structures from the analysis of domain-averaged fermi holes. Nature of the Mn-Mn bond in bis(pentacarbonylmanganese). J Comput Chem 26:447–454
87. Ponec R, Yuzhakov G, Carbó-Dorca R (2003) Chemical structures from the analysis of domain-averaged Fermi holes: multiple metal-metal bonding in transition metal compounds. J Comput Chem 24:1829–1838
88. Reinhold J, Kluge O, Mealli C (2007) Integration of electron density and molecular orbital techniques to reveal questionable bonds: the test case of the direct Fe−Fe bond in $Fe_2(CO)_9$. Inorg Chem 46:7142–7147
89. Overgaard J, Platts JA, Iversen BB (2009) Experimental and theoretical charge-density study of a tetranuclear cobalt carbonyl complex. Acta Crystallogr B 65:715–723
90. Blanco MA, Martín Pendás A, Francisco E (2005) Interacting quantum atoms: a correlated energy decomposition scheme based on the quantum theory of atoms in molecules. J Chem Theory Comput 1:1096–1109
91. Ponec R (1997) Electron pairing and chemical bonds. Chemical structure, valences and structural similarities from the analysis of the Fermi holes. J Math Chem 21:323–333
92. Ponec R, Roithová J (2001) Domain-averaged Fermi holes – a new means of visualization of chemical bonds. Bonding in hypervalent molecules. Theor Chem Acc 105:383–392
93. Ponec R, Duben AJ (1999) Electron pairing and chemical bonds: bonding in hypervalent molecules from analysis of Fermi Holes. J Comput Chem 20:760–771
94. Ponec R, Uhlik F (1997) Electron pairing and chemical bonds on the accuracy of the electron pair model of chemical bond. J Mol Struct (Theochem) 391:159–168
95. Savin A, Silvi B, Colonna F (1996) Topological analysis of the electron localization function applied to delocalized bonds. Can J Chem 74:1088–1096
96. Silvi B (2002) The synaptic order: a key concept to understand multicenter bonding. J Mol Struct 614:3–10

97. Bader RFW, Essén H (1984) The characterization of atomic interactions. J Chem Phys 80:1943–1960
98. Cremer D, Kraka E (1984) A description of the chemical bond in terms of local properties of electron density and energy. Croat Chem Acta 57:1259–1281
99. Abramov YA (1997) On the possibility of kinetic energy density evaluation from the experimental electron-density distribution. Acta Crystallogr A 53:264–272
100. Jayatilaka D, Grimwood DJ (2001) Wavefunctions derived from experiment. I. Motivation and theory. Acta Crystallogr A 57:76–86
101. Jayatilaka D, Grimwood DJ (2004) Electron localization functions obtained from X-ray constrained Hartree-Fock wavefunctions for molecular crystals of ammonia, urea and alloxan. Acta Crystallogr A 60:111–119
102. Espinosa E, Alkorta I et al (2002) From weak to strong interactions: a comprehensive analysis of the topological and energetic properties of the electron density distribution involving X-H···F-Y systems. J Chem Phys 117:5529–5542
103. Shriver DF, Atkins PW (1999) Inorganic chemistry, 3rd edn. Oxford University Press, Oxford
104. Overgaard J, Platts JA, Iversen BB (2009) Experimental and theoretical charge-density study of a tetranuclear cobalt carbonyl complex. Acta Crystallogr B 65:714–723
105. Green SP, Jones C, Stasch A (2007) Stable Magnesium(I) compounds with Mg-Mg bonds. Science 318:1754–1757
106. Resa I, Carmona E et al (2004) Decamethyldizincocene, a stable compound of Zn(I) with a Zn-Zn bond. Science 305:1136–1138
107. Dunitz JD (1995) Phase changes and chemical reactions in molecular crystals. Acta Crystallogr B 51:619–631
108. Brock CP, Fu Y (1997) Rigid-body disorder models for the high-temperature phase of Ferrocene. Acta Crystallogr B 53:928–938
109. Allen FR (2002) The Cambridge structural database: a quarter of a million crystal structures and rising. Acta Crystallogr B 58:380–388
110. Coriani S, Haaland A et al (2006) The equilibrium structure of Ferrocene. ChemPhysChem 7:245–249
111. Bader RFW, Matta CF, Cortés-Guzmán F (2004) Where to draw the line in defining a molecular structure. Organometallics 23:6253–6263
112. Bader RFW, Matta CF (2001) Bonding to Titanium. Inorg Chem 40:5603–5611
113. Kubas GJ (2001) Metal dihydrogen and σ-bond complexes. Kluwer Academic/Plenum, New York
114. Hoyano JK, Elder M et al (1969) Hydrogen-bridged silicon-rhenium bonds. A diphenylsilane complex of rhenium carbonyl. J Am Chem Soc 91:4568–4569
115. Geftakis S, Ball GE (1998) Direct observation of a transition metal alkane complex, $CpRe(CO)_2$(cyclopentane), using NMR spectroscopy. J Am Chem Soc 120:9953–9954
116. Schubert U (1990) η^2 Coordination of Si-H σ bonds to transition metals. Adv Organomet Chem 30:151–187
117. Lichtenberger DL, Rai-Chaudhuri A (1989) Electronic structure factors of silicon-hydrogen bond activation by transition metals. The valence photoelectron spectrum of silylmanganese complex $(\eta^5\text{-}C_5H_5)Mn(CO)_2HSiCl_3$. J Am Chem Soc 111:3583–3591
118. Lichtenberger DL, Rai-Chaudhuri A (1990) Electronic structure control of silicon-hydrogen bond activation by transition metals. 2. Valence photoelectron spectra of $(\eta^5\text{-}C_5H_4CH_3)Mn(CO)_2HSiPh_3$, $(\eta^5\text{-}C_5H_4CH_3)Mn(CO)_2HSiHPh_2$ and $(\eta^5\text{-}C_5H_4CH_3)Mn(CO)_2HSiFPh_2$ (Ph = C_6H_5). J Am Chem Soc 112:2492–2497
119. Schubert U, Ackermann K, Wörle B (1982) A long silicon-hydrogen bond or a short silicon-hydrogen nonbond? Neutron-diffraction study of $(\eta^5CH_3C_5H_4)(CO)_2(H)MnSiF(C_6H_5)_2$. J Am Chem Soc 104:7378–7380
120. Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Can J Phys 58:1200–1211

121. Valenzano L, Torres FJ et al (2006) Ab initio study of the vibrational spectrum and related properties of crystalline compounds; the case of $CaCO_3$ Calcite. Z Phys Chem 220:893–912
122. Dunning TH (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J Chem Phys 90:1007–1023
123. Muenter JS (1975) Electric dipole moment of carbon monoxide. J Mol Spectrosc 55:490–491
124. Hehre WJ, Radom L et al (1986) Ab initio molecular orbital theory. Wiley, New York
125. Bader RFW, Matta CF (2004) Atomic charges are measurable quantum expectation values: a rebuttal of criticisms of QTAIM charges. J Phys Chem A 108:8385–8394
126. Gillet J-M (2007) Determination of a one-electron reduced density matrix using a coupled pseudo-atom model and a set of complementary scattering data. Acta Crystallogr A 63:234–238
127. Cade PE, Bader RFW et al (1971) Molecular charge distributions and chemical bonding. V. Molecular excitation, ionization and electron attachment. J Chem Phys 54:3517–3533
128. Francisco E, Martín Pendás A, Blanco MA (2009) A connection between domain-averaged Fermi hole orbitals and electron number distribution functions in the real space. J Chem Phys 131:124125

Index

A
2-Acetyl-1,8-dihydroxy-3,6-dimethylnaphthalene, 227
AMCHA, 47, 63
Amino acids, 36, 37
Ammonia borane, 134
Anharmonic nuclear motion, 185
Atomic displacement parameters (ADPs), 28
Atomicity, 3
Atoms in molecules (AIM) theory, 131

B
Bader's atoms in molecules (AIM) theory, 35, 131
Basis set effects, 47
Benzene, 214
Benzoylacetone, 224
Betaine, 224
BH_3CO, 276
Bond critical points (BCPs), 49, 196
Bonding descriptors, local/nonlocal, 193
Bragg structure factor, 3
Bullvalene trisepoxide, 179

C
Cambridge Structural Database (CSD), 51, 255
Cathepsin B, 47, 74
Charge density, 143
Chemical bonding/transferability, 193
CI-Singles-Doubles (CISD), 105
Cluster calculations, 35
Co–Co mp/bcp, 242
Coherent elastic diffraction, 2
Compensatory transferability, 205
Complete-active space self-consistent-field approach (CASSCF), 105
Conceptual DFT, 99, 128
Conductor-like screening model (COSMO), 62
Configuration state functions (CSFs), 104
Contact densities, 99, 125
Continuity equation, 99
Core/valence decompositions, 273
Current-density functional theory, 115
Cyclohexene/cyclohexadiene, 214
Cysteine proteases, inhibitors, 74

D
Data resolution, 31
Decamethyldizincocene, 253
Deformation densities, 230
Delocalization indices, 278
Density functional theory (DFT), 48, 99, 101, 115, 128
  conceptual, 128
Diborane, 211
Dipole moments, 32
  enhancements, 27, 37
Dirac–Coulomb–Breit (DCB) Hamiltonian, 99, 107, 113
Dirac hole theory, 109
Domain-averaged Fermi hole (DAFH) analysis 233
Douglas–Kroll–Hess (DKH), 99, 109

E
E64c, 47
Effective core potentials (ECPs), 129
Electron conjugation, 193

Electron correlation effects, 99
Electron density, 1ff, 3
Electron density distribution (EDD), 28, 102
Electron leak problem, 134
Electrostatic potentials (ESPs), 62
Embedding scheme, 131
Environmental effects, 47
Enzymes, 61
Ethane, 206
Extinction, 179
  coefficient, 145

F
$Fe_2(CO)_9$, 233
$Fe(CO)_5$, 276
$Fe(NO)^{2+}$, 120
Ferrocene, 255
Fractals, 155
Frozen-density embedding, 99, 102
Full configuration interaction (FCI), 104

G
α-Glycine, 10, 20, 177
Gross residual electrons, 156

H
Hamiltonian operator, 103, 106
Hansen–Coppens formalism, 5, 28
Hard and soft acid and base (HSAB) principle, 128
Hartree–Fock (HF) atomic wave functions, 5
HC pseudoatom model (HC-PA), 5
Heptasulfur imide, 7
Hirshfeld refinement, 27
Hirshfeld's stockolder partitioning, 202
Hirshfeld surfaces, 202
Hydrides, diatomic, 216
Hydrocarbons, 206, 210
  aromatic, 212
Hydrogen bonds (HB), SF description, 218
  short-strong, 193
Hydrogen scattering, 30

I
Imidazole, 224
Independent atom model (IAM), 38, 203
Interacting quantum atoms (IQA) theory, 233
Interaction densities, 230
Interatomic surfaces (IAS), 131
Invariom modelling, 38

K
Kohn–Sham equations with constraint electron density (KSCED), 133

L
Li–X, 207
Local excess electrons, 156
Local source (LS) profiles, 230
Low-barrier hydrogen bonds (LBHB), 219
Loxistatin, 63, 81
Loxistatinic acid, 63

M
Malonaldeyde, 222
Maximum entropy method (MEM), 176
Metal–ligand bonds, 193
Metal–ligands (M–L) interactions, 231
Metallocenes, 255
Metal–metal (M–M) bonds, 193, 231
Metal–silane σ-complexes, 254
Methylene-bis(triimido)sulfonic acid, 52
Methyl(diimido)sulfinic acid, 52
$M_2(formamidinate)_4$, 246
Mg–Mg bond, 253
$Mn_2(CO)_{10}$, 237
Molecular dipole moment, 27
Mulliken electronegativity, 128
Mulliken's population analysis, 269
Multipole expansion, 4
Multipole model, 27, 263
  refinement, 174, 263
Multipole-modeled primary density (MMPD), 263

N
Natural bond orbital/natural resonance theory (NBO/NRT), 54
Net residual electrons, 156
Nitromalonamide, 224
N-phenylpyrrole (PP), 181

O
Organometallics, metal–metal/metal–ligand bonds, 231

P
Perfect transferability, 205
Picric acid, 224

Picture change error (PCE), 99, 109
Polar bonds, 47
Population analysis, 193
Primary density (PD), 263
Prior derived F-constraints (PDCs), 176
Propane, 206
Pseudoatom density, absolute error, 10
Pseudoatom model/formalism, 5

Q
Quantum electrodynamical (QED) corrections, 107
Quantum mechanics/molecular mechanics (QM/MM), 35, 47, 62, 83
Quantum theory of atoms in molecules (QTAIM), 35, 48, 196

R
Radial functions (RDF), 1, 2
Rational drug design, 61
Real spherical harmonics (RSH), 2
Relativistic effects, 99
Residual density analysis (RDA), 143
Resonance-assisted hydrogen bonds (RAHBs), 219
Ring critical points (rcp), 203
Roby–Davidson partitioning, 73
R-values, 146

S
$S(N^tBu)_3$, 164
Scalar-relativistic one-component methods, 111
Short strong hydrogen bonds (SSHB), 218
Single-crystal X-ray diffraction, 27
Slater–Condon rules, 112
Slater determinant, 104, 113
Source function, 47, 83, 193, 195, 200
Stockholder partitioning, 1, 8
Stockholder pseudoatom databank, 18
Sulfurdiimide/sulfurtriimide, 52
Systematic sources of error, 143

T
Terbogrel, 19
Thermal diffuse scattering (TDS), 151
Titanium tetrachloride, 134
Topology, 143
Total error, 156
Transferability, 205
Transition metal complexes, 121, 254
Trimethylenemethane (TMM), 255

U
Unrestricted stockholder pseudoatoms (U-SPA), 9

V
Valence-shell charge concentrations (VSCCs), 49, 153

W
Wave function, 103
  fitting, 27, 40

X
X-ray diffraction, 1

Z
Zero-flux surfaces, 131
Zn(I)–Zn(I) single bond, 253
ZORA, 99, 111
ZORA-4 density, 114